Functional analysis of the competence transcription factor ComK of Bacillus subtilis

Susanna, Kim Amé

Functional analysis of the competence transcription factor ComK of *Bacillus subtilis*

Functionele analyse van de competentie transcriptie factor ComK van *Bacillus subtilis*

Rijksuniversiteit Groningen

Functional analysis of the competence transcription factor ComK of *Bacillus subtilis*

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, Dr. F. Zwarts,
in het openbaar te verdedigen op
vrijdag 27 mei 2005
om 14.45 uur

door

Kim Amé Susanna

geboren op 3 juli 1976
te Heerenveen

Promotor:	Prof. Dr. O. P. Kuipers
Copromotor:	Dr. L. W. Hamoen
Beoordelingscommissie:	Prof. Dr. J. M. van Dijl
	Prof. Dr. P. J. M. van Haastert
	Prof. Dr. I. J. van der Klei

If you don’t know where you’re going,
any road will take you there
(George Harrison, 1943 – 2001)

Paranimfen:
Chris den Hengst
Esther de Jong

Support
This Ph. D. study was carried out in the Groningen Biomolecular Sciences and Biotechnology Institute (Faculty of Mathematics and Natural Sciences, University of Groningen).

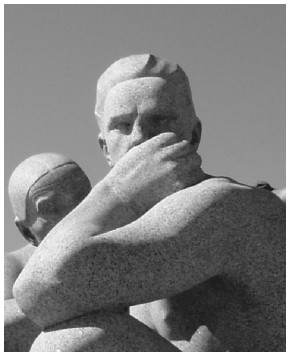
Printed by
This thesis was printed by 'Het Facilitair Bedrijf', University of Groningen, The Netherlands.

Cover
The picture on the cover is taken at the Haukelifjell in Norway.

Contents

Chapter 1

General introduction



Bacillus subtilis, a Gram-positive model bacterium

Bacillus subtilis is a rod-shaped Gram-positive bacterium, which belongs to the family of Bacillaceae (fig. 1). This family of spore-forming bacteria consists of different genera, which show a wide variety in morphology, ranging from cocci and rods to mycelial bacteria. *B. subtilis* is placed into the *Bacillus* genus. Grouping of species into this genus was historically mainly based on morphological and physiological characteristics, like the shape of the cells and the requirement of oxygen for growth (Priest, 1993). Nowadays, the availability of a large number of complete genome sequences, enables a more precise and better classification of bacterial species.

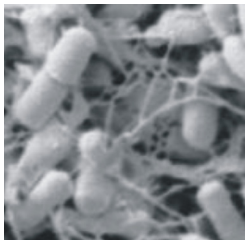


Figure 1. Electron microscopy picture of *B. subtilis* (www.magma/ca/~scimat)

The species of the *Bacillus* genus live in the soil, but can also be found in associated environments, like plants and plant-materials, marine and fresh-water surroundings and in food and animals. The genus contains extremophiles as well as non-extremophiles, pathogens and non-pathogens. In addition to *B. subtilis*, other well-known members of the *Bacillus* genus are the industrially important species *B. licheniformis, B. amyloliquefaciens* and *B. stearothermophilus*, which are widely used for the production of both native and heterologous proteins (Harwood, 1992), and *B. subtilis* (*natto*), which is used in food industry for the production of Natto, a traditional Japanese fermented soybean product. Furthermore, the genus contains pathogenic species like the food-spoiling *B. cereus* and the notorious *B. anthracis*, the etiological agent of anthrax (review: Mock and Fouet, 2001). *B. subtilis* is generally regarded as the model-organism of the *Bacillus* genus, and in fact of all Gram-positive bacteria, because of its long research history, which characterized the species as a genetically easy accessible GRAS (generally regarded as safe)-organism (Harwood, 1992; Priest, 1993).

History of the *B. subtilis* research

Since the first clear characterization of the Marburg strain as *Bacillus subtilis* in 1936, the species has been extensively studied. Three different pioneer research-lines can be distinguished in the mid-20th century. i) Random mutagenesis with UV-light and X-rays provided a source of auxotrophic mutants. The genetics of several biochemical pathways were studied and these mutants were used for the detailed investigation of genetic recombination (Burkholder and Giles, 1947). ii) Other pioneer research was performed by Anagnostopoulos and Spizizen (1961), who optimized the transformation efficiency, resulting in the method that is nowadays still used in almost all *B. subtilis* research. (iii) The third research-line consisted of the investigation of sporulation, starting with the isolation and mapping of sporulation mutants. At the same time, the order of morphological changes during the sporulation process was determined using electron microscopy (Hoch, 1971; Piggot and Coote, 1976; Ryter *et al*., 1966).

B. subtilis *sequencing project*

A major breakthrough in *B. subtilis* research was achieved in 1997, when the combined effort of 34 groups resulted in the completion of the *B. subtilis* 168 genome sequence (Kunst *et al.*, 1997). The complete genome consists of a 4.2 Mb single chromosome with an average G+C content of 43.5%. In total, 4107 open reading frames were predicted, of which 303 are located on prophages. For about 1500 of the genes, a (putative) function could be assigned, partially based on performed research and partially based on sequence homology.

As a follow-up of the sequencing project, the *Bacillus subtilis* gene function analysis (BSFA) project aimed for the elucidation of the role of the remaining genes with unknown function, the so-called *y*-genes. Of the more than 2500 genes for which disruption mutants have been constructed, only 271 genes were shown to be essential for viability of *B. subtilis* under laboratory conditions. This relatively low number may well be an underestimate, since *B. subtilis* contains a large number of gene paralogues. Most of the essential genes could be classified in the group of cell metabolism genes, involved in information processing, synthesis of the cell envelope, determination of cell shape, cell division and cell energetics (Kobayashi *et al*., 2003).

Adaptation and differentiation processes in *B. subtilis*

From the early years of research on, *B. subtilis* is well-known for its ability to adapt to varying conditions in its environment, one of the most fascinating aspects of the species. Under optimal growth conditions, cells divide every twenty to thirty minutes, resulting in an exponential growth of the culture. When nutrient availability becomes limiting, at the end of the exponential growth-phase, *B. subtilis* reacts by initiating a sequence of adaptation processes (fig. 2; review: Msadek, 1999), starting with antibiotic production and cell motility (fig. 2, nrs. 1 and 2). The complex motility and chemotaxis pathway enables the cell to swim, thereby increasing its chance to enter a more favourable environment with respect to the presence or absence of competitor species and/or nutrients (Aizawa *et al*., 2002), while antibiotic production provides the cell with the means to eliminate other organisms, competing for the same, limited pool of nutrients (Zuber *et al*., 1993). A prolonged limitation of nutrient availability results in entry into the stationary growth-phase, characterized by a reduction of growth of the population. In the search for available nutrients, *B. subtilis* cells initiate the production and secretion of degradative enzymes, like proteases (fig. 2, nr. 3), which enables the use of a wider variety of nutritional sources (Kunst *et al*., 1994).

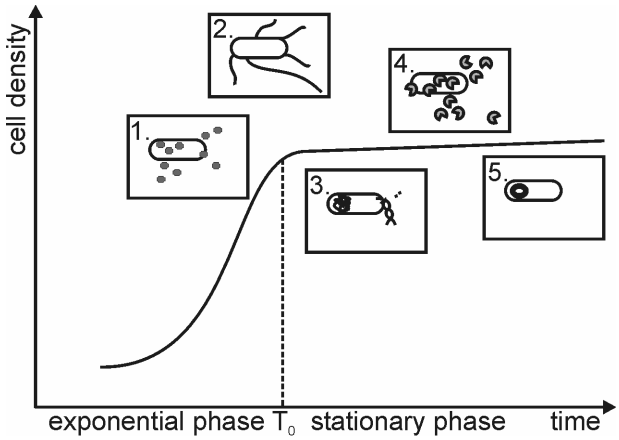


Figure 2. Representation of growth-stage related adaptations in *B. subtilis*
T_0 marks the transition point from the exponential to the stationary growth-phase. The adaptation and differentiation processes are initiated at the end of the exponential growth-phase with antibiotic production and cell motility. The indicated numbers refer to the descriptions in the text.

When nutritional stress conditions continue, *B. subtilis* activates two more drastic differentiation processes: genetic competence (fig. 2, nr. 4; discussed below) and, ultimately, sporulation (fig. 2, nr. 5). Competent cells are able to take up macromolecular DNA, which can be integrated into the genome, thereby providing the cell with new genetic abilities (review: Dubnau and Lovett, 2002). Alternatively, the internalised DNA could be used as template for DNA-repair (Michod *et al*., 1988) and single nucleotides could serve as food source (Redfield *et al*., 1997). However, the latter two possibilities are not considered as the main reasons for cells to develop competence (Dubnau, 1999). The last process, sporulation, results in the formation of highly resistant dormant spores, which are able to survive harsh conditions for a long time and can finally germinate to start a new *B. subtilis* population (review: Hilbert and Piggot, 2004).

Competence for transformation, an intriguing process

Genetic or natural competence is a differentiation process, which enables *B. subtilis* to take up exogenous DNA and to incorporate it into its genome in order to acquire new genetic abilities, which may help to survive harsh conditions. In addition to *B. subtilis*, natural competence has been observed in a wide variety of Gram-negative (*e. g. Haemophilus influenzae*, *Neisseria gonorrhoeae*) and Gram-positive bacteria (*e. g. Streptococcus pneumoniae*) and even Archaea (*e. g Methanococcus voltae*) (Dubnau, 1999; Lorenz and Wackernagel, 1994).
The most intriguing observation about competence development in *B. subtilis* is the heterogeneity of the process. Even when all conditions are optimal, competence development is limited to a maximum of about 10% of the population, although all cells share the same genetic background (Hadden and Nester, 1968; Haseltine-Cahn and Fox, 1968; Nester and Stocker, 1963). An attractive speculation is that this limitation is a sort of safety mechanism to increase the chance that at least part of the culture will survive unfavourable conditions. Some of the cells might survive through competence development, while others could survive by initiating another differentiation process, *e. g.* sporulation, which is also heterogeneous and, likely, mutually exclusive with competence (Chung *et al*., 1994).

Changes in cell physiology of competent cells
Differentiation into competent cells is accompanied by several drastic changes in cell physiology, like a block of DNA replication, cell-wall synthesis and cell division (Dooley *et al.*, 1971; Haijema *et al*, 2001; Nester and Stocker, 1963), although the appearance of competent cells resembles that of non-competent stationary growth-phase cells in occurring mainly as single cells with single nucleoids and rarely containing FtsZ-rings (Haijema *et al*., 2001). Competent and non-competent cells can be distinguished based on differences in their buoyant density, which is higher for competent than for non-competent cells. These characteristics can be used to separate the subpopulations by density gradient ultracentrifugation (Hadden and Nester, 1968; Haijema *et al*., 2001; Haseltine-Cahn and Fox, 1968). Using this technique, it was demonstrated that the major players in competence development are expressed only in the competent cell fraction (Albano *et al*., 1987; Hahn *et al*., 1994).

Requirements for the development of competence
Competence development in *B. subtilis* is initiated only under specific conditions in a growth-phase dependent manner. Under laboratory conditions, competence develops in minimal medium with glucose as the sole carbon source. Differentiation into competent cells is started at the onset of the stationary growth-phase and, in addition to low nutrient availabilities, requires high cell densities. An optimum for natural competence is reached after two hours into the stationary growth-phase (Dubnau and Lovett, 2002; Solomon and Grossman, 1996).
The regulatory pathway leading to competence development in *B. subtilis* will be discussed in detail below. In short, differentiation into competent cells depends on the activity of a key regulator, the competence transcription factor ComK. Competence regulation involves mainly the control of ComK-levels in the cell. ComK acts as a transcriptional activator, which first activates transcription of its own gene. Autostimulation results in a rapid increase of ComK-concentrations, followed by the activation by ComK of the late competence genes, which encode the DNA-binding, -uptake and -recombination machinery. Ultimately, assembly of this machinery enables the cell to become competent for transformation (reviews: Dubnau and Lovett, 2002; Hamoen *et al*., 2003b; Solomon and Grossman, 1996; Tortosa and Dubnau, 1999).

The DNA-binding, -uptake and recombination machinery
DNA-binding and -uptake occur at specific sites on the surface of the cell. *B. subtilis* is estimated to contain about 50 of these DNA-binding sites per competent cell (Dubnau and Cirigliano, 1972b; Singh, 1972). Early research showed that i) there is no sequence specificity for DNA-binding, ii) Double-stranded DNA binds to the cell and, after fragmentation, single-stranded DNA fragments with sizes up to 20 kb are taken up, iii) DNA-uptake is a rapid process: only minutes after binding, one strand is transported into the cell, while the non-transported strand is degraded into nucleotides and released into the medium and iv) *B. subtilis* can internalize DNA from plasmid, phage or chromosomal origin (Contente and Dubnau, 1979; Davidoff-Abelson and Dubnau, 1973; Dubnau and Cirigliano, 1972a; Levine and Strauss, 1965; Piechowska and Fox, 1971; review: Dubnau, 1999).

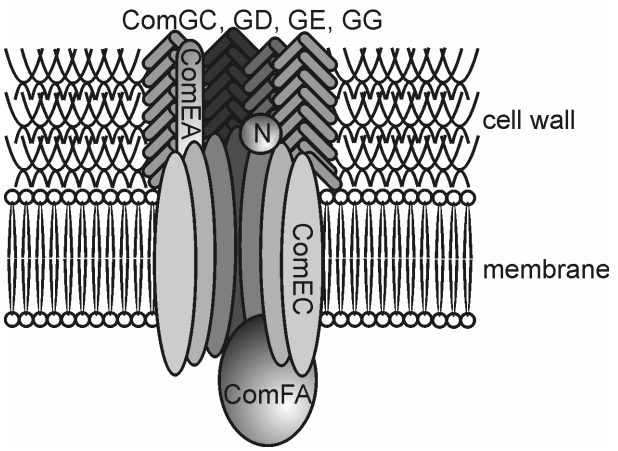


Figure 3. Overview of the DNA-binding and –uptake machinery
The ComG-proteins provide access through the cell-wall to ComEA, which binds double-stranded DNA. DNA is internalized in single-stranded form via a channel across the membrane formed by ComEC. The non-transported strand is degraded by NucA (N). The required energy for the transport is provided by ComFA. For details, see text (Figure adapted from Dubnau, 1999).

Each DNA-binding and -uptake site consists of a complex of proteins (fig. 3), encoded by the late competence genes *comC*, *comE*, *comF*, *comG and nucA* (review: Dubnau and Lovett, 2002). The first step of DNA-uptake is mediated by the seven proteins encoded by the *comG*-operon, which are all individually required

for DNA-binding and -uptake. Four of these proteins, ComGC, GD, GE and GG, share homology with type IV pilin-like proteins. The peptidase ComC is involved in protein-processing, required for the correct assembly of the ComG-proteins into a pilin-like structure. This structure provides access of the exogenous, double-stranded DNA through the cell-wall rather than binding DNA itself (Chung and Dubnau, 1995 and 1998; Chung *et al*., 1998). Once the DNA has passed the cell-wall, it associates with ComEA, which is a membrane-spanning, cell-wall associated protein functioning as a DNA-receptor by binding dsDNA to its C-terminus (Inamine and Dubnau, 1995; Provvedi and Dubnau, 1999). In addition to DNA-binding, ComEA is required for transport of single-stranded DNA across the membrane through an aqueous channel formed by the pore-forming protein ComEC (Inamine and Dubnau, 1995). Internalization of ssDNA into the cell is accompanied by degradation of the other DNA-strand into medium-released nucleotides by the membrane-localized nuclease NucA (Provvedi *et al*., 2001). Likely, the motor of the transport process is ComFA, which resembles a DNA-helicase and is located on the inner side of the membrane. ComFA contains a single ATP-binding site and has been suggested to function as an ATPase, providing the energy to drive the DNA-transport (Londono-Vallejo and Dubnau, 1993; 1994a and b). As soon as the new DNA is inside the cell, a heteroduplex is formed between the exogenous ssDNA and the helix of the genomic DNA of the cell (Bodmer and Ganesan, 1964). Integration of the ssDNA occurs via homologous recombination, catalysed by the recombination protein RecA and the DNA-helicase AddAB (Haijema *et al*., 1996; review: Dubnau and Lovett, 2002).

Regulation of competence development

Activation of the competence machinery depends on the presence and activity of the competence transcription factor ComK in the cell. The ComK-level can rapidly increase via an autostimulatory loop. Once the concentrations are sufficiently high, competence is inevitable. In order to prevent premature *comK*-expression, which could be fatal for the cells, *B. subtilis* acquired a complex regulatory pathway (fig. 4), consisting of both transcriptional and post-translational control of ComK (Dubnau and Lovett, 2002; Hamoen *et al*., 2003b; Solomon and Grossman, 1996).

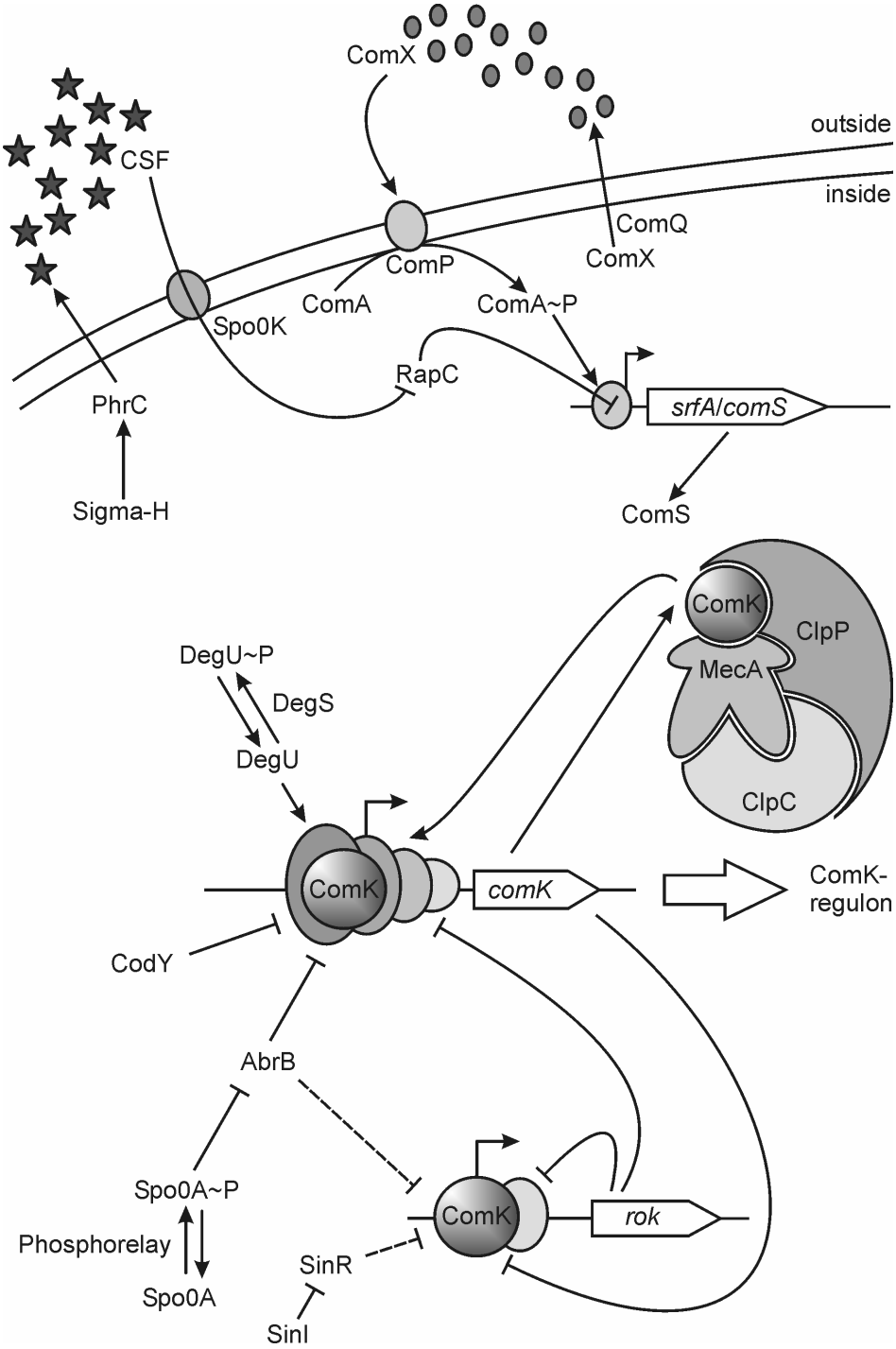

ComX
outside
inside
CSF
ComQ
ComX
ComP
ComA
ComA~P
Spo0K
RapC
PhrC
srfA/comS
Sigma-H
ComS
ComK
ClpP
MecA
ClpC
DegU~P
DegS
DegU
ComK
comK
ComK-
regulon
CodY
AbrB
Spo0A~P
Phosphorelay
Spo0A
SinR
SinI
ComK
rok

Figure 4, previous page. Schematic overview of the regulatory pathway controlling competence development in *B. subtilis*
Positive and negative regulations are indicated by arrows and perpendiculars, respectively. Dotted lines represent putative regulation. The *comK*, *srfA* and *rok* promoters are shown to indicate transcriptional regulation. Pheromone transport across the membrane is shown. The major regulatory proteins are indicated with symbols. For details, see text (Figure adapted from Hamoen *et al*., 2003b)

Transcriptional repression of comK

During the exponential growth-phase, *comK*-transcription is repressed by binding of the repressor proteins AbrB, CodY and Rok to the *comK*-promoter. AbrB is a transition-state regulator, which is involved in growth-phase dependent regulation of different stationary phase processes (Strauch and Hoch, 1993). It is a small DNA-binding protein, which binds to the *comK*-promoter in the region between -80 and +40, relative to the transcription start. By occupying the -35 and -10 sequences, AbrB prevents transcription (Hamoen *et al*, 2003a). At the onset of stationary growth, *abrB*-transcription is repressed by Spo0A~P, AbrB-levels decrease and *comK*-repression is relieved (Hahn *et al*., 1995).
The response of *comK*-expression to nutrient limitation is mediated by the nutritional repressor CodY, a GTP-binding transcription factor which senses the intracellular GTP-concentration as an indication for the nutritional state of the cell. During the exponential growth-phase, with an excess of nutrients, the GTP-concentration is high and GTP-bound-CodY represses transcription of *comK*, by binding to the *comK*-promoter, partially overlapping with the RNA polymerase binding site. In addition to competence, CodY inhibits a wide variety of stationary phase genes. Upon conditions of nutrient deprivation, the GTP-concentration drops and repression by CodY is relieved, allowing the controlled stationary phase genes to be expressed (Ratnayake-Lecamwasam *et al*., 2001; Serror and Sonenshein, 1996; Slack *et al.*, 1995).
Only recently, a third repressor of *comK*-transcription was discovered, *i. e.* Rok (repressor of *comK*-expression), a DNA-binding protein that acts directly on the *comK*-promoter (Hoa *et al*., 2002). Rok and ComK can bind simultaneously to the *comK*-promoter as well as to the *rok*-promoter, thereby establishing a complex regulatory feedback loop (Hoa *et al*., 2002; W. K. Smits, unpublished results). Autorepression of *rok* most likely serves to prevent Rok-overexpression, which is probably lethal to the cell

(Hoa *et al*., 2002). ComK-controlled *rok*-transcription establishes a second positive feedback loop involved in competence regulation. When intracellular ComK-levels increase, ComK and Rok repress expression of *rok*, thereby enhancing the autostimulatory effect of ComK on its own expression to ensure a rapid increase in ComK-production in the cell (Hoa *et al.*, 2002).
The discovery of Rok as a repressor of ComK provided an explanation for the ill-understood positive requirements for AbrB and SinR in competence development (Hahn *et al*., 1995 and 1996). Like AbrB, SinR is a transition state regulator, which is involved in the regulation of several stationary growth processes. Previously, it was shown that there is little or no *comK*-expression in *sinR* or *abrB* deletion strains (Hahn *et al*., 1996). Especially the role of AbrB has long been unclear, since it acts as an activator as well as a repressor of competence. The dual role was shown to depend on the AbrB-concentration in the cell, which should be within a certain window in order to stimulate competence development (Hahn *et al*., 1995). Both SinR and AbrB were demonstrated to downregulate *rok*-transcription, either direct or indirect. The repressing effect of both regulators on *rok*-transcription, results in indirect stimulation of expression of *comK* (Hoa *et al.*, 2002).

Post-translational control of ComK
Although *comK*-expression is tightly regulated during exponential growth by transcriptional repression, a basal level of ComK is produced in this stage. To prevent this ComK from activating the late competence genes, and thereby initiating a premature development of competence, it is essential to keep ComK inactive until the transition to the stationary growth-phase. For this purpose, *B. subtilis* uses a post-translational control system (Kong and Dubnau, 1994; Msadek *et al*., 1994; Turgay *et al*., 1997, 1998). In the exponential growth-phase, any produced ComK is trapped in a protein complex by binding to the N-terminus of the adaptor protein MecA. The C-terminus of MecA interacts with the ATP-dependent protease ClpC, thereby targeting both ComK and MecA for proteolytic degradation by ClpP (Persuh *et al*., 1999). Loss of function mutations in either MecA or ClpC result in a hyper-competent phenotype. In these strains, ComK is expressed in all cells in a medium-independent manner, so competence can develop even in rich media. However, expression of ComK is still growth-phase dependent (Dubnau and Roggiani, 1990).

Quorum-sensing regulation of competence development
Competence develops in response to changes in environmental conditions, like nutrient deprivation and increased cell densities in the medium. To provoke intracellular responses on extracellular events, environmental signals should be sensed and interpreted by the cell in order to initiate the proper responses. For this purpose, many bacteria make use of so-called quorum-sensing pathways in which a two-component system reacts on environmental stimuli by phosphorylation, and thereby activation, of a cytoplasmic protein, which in turn is involved in the initiation of an intracellular response (Perego and Hoch, 2002). To establish the initiation of competence development upon increasing cell densities, *B. subtilis* cells produce quorum-sensing pheromones during the exponential growth-phase. These small peptides are secreted into the medium, where they act in cell-cell communication. In time, the pheromone concentration in the medium accumulates, as an indication for an increase in cell density. In the case of competence development, two quorum-sensing peptides are involved: ComX and CSF (Magnuson *et al*., 1994; Solomon *et al*., 1996).
ComX, the main competence stimulating pheromone, is synthesized as an inactive precursor in the cytosol. Activation of ComX requires cleavage to a decapeptide with an isoprenyl modification of the tryptophane residue. This process, as well as the secretion of ComX, requires ComQ. The *comQ* and *comX* genes are located in an operon together with the genes encoding the two-component system ComP/ComA (Ansaldi *et al.*, 2002; Magnuson *et al*., 1994; Weinrauch *et al.*, 1991). Accumulation of ComX in the medium is sensed by the sensor-kinase ComP, which responds by autophosphorylation and, subsequently, by phosphorylation of the response-regulator ComA (Solomon *et al*., 1995). ComA~P acts as a transcriptional activator at the promoter of the *srfA*-operon, leading to the production of the lipopeptide antibiotic SrfA (Nakano and Zuber, 1989; Roggiani and Dubnau, 1993). Simultaneously, the *comS* gene, which lies embedded in the *srfA*-operon, is transcribed, resulting in the synthesis of ComS (D'Souza *et al.*, 1994; Hamoen *et al*., 1995). ComS is a small protein, which is essential for competence development because of its role in the relief of post-translational control of ComK. ComS can liberate ComK from the complex with MecA and ClpCP by destabilizing this complex. ComS binds to the N-terminus of MecA, thereby replacing ComK, which is then released from the complex and escapes

proteolytic degradation. In turn, the MecA-ClpC interaction now targets ComS and MecA for proteolysis by ClpP (Persuh *et al*., 1999; Turgay *et al.*, 1997 and 1998).
The second pheromone, the competence stimulating factor (CSF), is important for modulation of the response to ComX-accumulation. CSF is produced as part of the fourty amino acids long PhrC-protein and requires cleavage to a pentapeptide to become active (Kanamaru *et al*., 2002; Solomon *et al*., 1996; Stephenson *et al*., 2003). Secreted CSF is transported back into the cell by the oligopeptide permease Spo0K and, depending on the concentration, it activates or inhibits different adaptation processes. At lower concentrations, CSF stimulates competence, while at high range concentrations, it is involved in initiation of sporulation and inhibition of competence, thereby defining a time-window for competence development (Lazazzera *et al*., 1997). CSF acts on competence via interaction with RapC, which can, in the absence of CSF, inhibit binding of ComA~P to the *srfA*-promoter. By binding to RapC, CSF prevents RapC from inhibiting ComA~P-DNA binding, thereby indirectly enhancing *srfA*- and thus *comS*-transcription activation by ComA~P (Core and Perego, 2003; Solomon *et al*., 1996).

Transcriptional activation at the comK-*promoter*
Competence development is rapidly proceeding once ComK is released from the proteolytic complex. ComK functions as the key regulatory protein in competence development by activating gene transcription via binding to specific ComK-binding sites, so-called K-boxes, located upstream of competence genes, including *comK* itself (Hamoen *et al*., 1998; Van Sinderen and Venema, 1994b; Van Sinderen *et al*., 1995). As will be discussed in more detail below, there are three classes of K-boxes, each containing two sites for ComK-binding, separated by a discrete number of helical turns (Hamoen *et al*., 1998). The K-box of the *comK*-promoter displays the largest distance between the two ComK-binding sites and is therefore of the least efficient type for both ComK-binding and transcription activation, especially when ComK-concentrations are low. In the early stages of *comK*-expression, transcription activation at the *comK*-promoter therefore requires the help of DegU. This priming protein binds the DNA at the spacing region between the two ComK-binding sites of the K-box, thereby strongly stimulating transcription. The autostimulatory circuit involved in *comK*-transcription results in a boost of ComK-production and, thereby,

a bypass of the requirement for DegU (Hamoen *et al.*, 2000). The rapid increase in free-ComK-levels in the cell, functions as an important enhancer of competence development. In addition to the transcription of its own gene, ComK stimulates transcription of other genes in the K-regulon, *e. g.* the late competence genes. Synthesis of the gene products, followed by the assembly of the different components, provides the cell with a functional DNA-binding, -uptake and -integration machinery, making the cell competent for transformation (Dubnau and Lovett, 2002).

Escape from competence
Once the growth conditions in the bacterial population's environment improve, the competent subpopulation has to escape competence to resume normal growth. During stationary growth, *B. subtilis* cells display a phenotype of single cells with low metabolic activity. Competent cells maintain this appearance for about two to three hours after the improvement of growth conditions, while non-competent cells rapidly start growth activities, like chaining, nuclear division and FtsZ-ring formation. In 2001, Haijema *et al.* revealed an additional role for ComGA in delaying resumption of growth in competent cells. Besides being required for DNA-binding and -uptake, ComGA is involved in preventing growth, DNA-replication and cell division until the remaining ComK in the cell is degraded by proteolysis.

ComK, the master regulator of competence development

ComK-regulated genes are characterized by the presence of a specific ComK-binding site, a K-box, upstream of their promoter (fig. 5). Extensive studies by Hamoen *et al*. (1998) revealed the complex nature of K-boxes. Functional ComK is thought to act as a tetramer, composed of two dimers. In each K-box, two binding sites for a ComK-dimer, the so-called AT-boxes, can be discerned (fig. 5A). AT-boxes are named after the consensus sequence consisting of a four basepair A-stretch and a four basepair T-stretch separated by five arbitrary nucleotides (AAAA-$(N)_5$-TTTT). Based on the length of the spacing region between both AT-boxes, three classes of K-boxes can be distinguished: Class I, II and III, with a spacing of 8, 18 or 31 nucleotides, respectively. The length of the spacing results in an interval of two, three or four complete heli-

cal turns, respectively, calculated from the starting A of the first AT-box to the starting A of the second AT-box (fig. 5B and C). In this way, both AT-boxes are located on the same side of the DNA-helix. This implies that also both bound ComK-dimers are positioned on one side of the helix, enabling interactions between the dimers to form a tetramer. Representatives of each K-box class are *addAB*, *recA* and *nucA* (Class I), *comC*, *comE*, *comF* and *comG* (Class II) and *comK* (Class III).

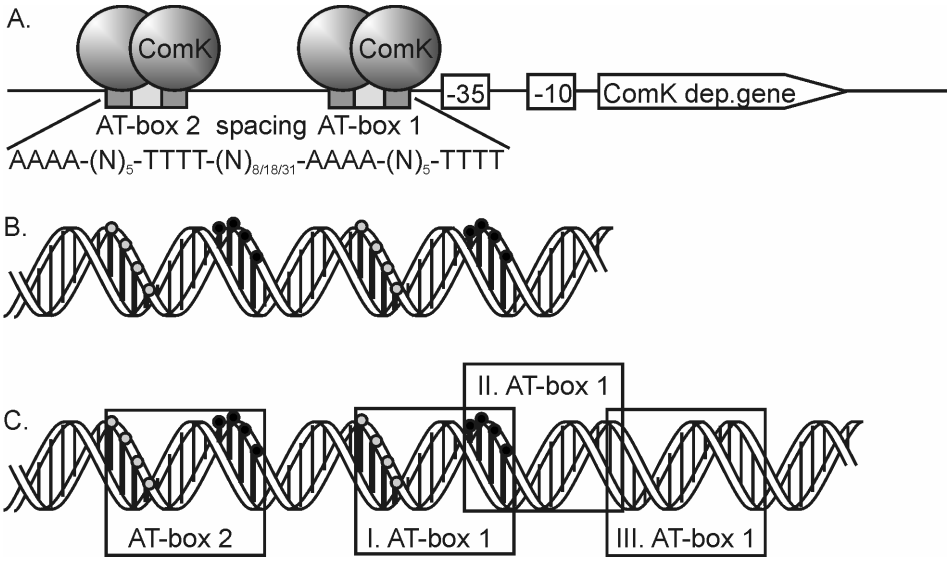


Figure 5. Overview of a K-box
(A) A K-box consists of two AT-boxes, separated by a flexible spacing, and is located upstream of a ComK-regulated gene. ComK binds to a K-box as a tetramer consisting of two dimers. Each ComK-dimer binds to an AT-box. (B) Helical projection of a K-box. A's are represented by grey circles and T's by black circles. (C) Overview of the three classes of K-boxes. I, II or III AT-box 1 indicates the position of this AT-box for each class relative to a fixed AT-box 2 (Figure adapted from Hamoen *et al*., 1998).

Occurrence of K-boxes in the B. subtilis *genome*

In the AT-rich genome of *B.* subtilis, the presence of a K-box is not a rare event. Previous research demonstrated that K-boxes with up to three deviations from the consensus sequence could still be functional targets for activation of transcription by ComK

(Hamoen *et al*., 1998). *In silico* analyses of the *B. subtilis* 168 genome revealed the presence of 1062 K-boxes, matching those criteria. About one-third of these boxes are located in intergenic regions. Only eight percent of the boxes were shown to be activated by ComK under laboratory conditions (Hamoen *et* al., 2002), suggesting that other criteria than the sole presence of a K-box are involved in determining whether or not a gene is ComK-regulated and, when regulated, to which level it is activated.

Characteristics of ComK-binding
As can be concluded from the variation in spacing length between the three classes of K-boxes, the DNA-binding abilities of ComK should be flexible as well. In addition to this flexibility, the binding behaviour of ComK to the K-boxes displays other interesting characteristics. For example, ComK-dimers bind the DNA at the AT-boxes in a sequence-specific way by establishing contacts through the minor groove of the DNA-helix (Hamoen *et al*., 1998). This behaviour is somewhat extraordinary, because, in general, protein-DNA interactions via the minor groove are considered as less sequence-specific than interactions through the major groove of the DNA, since the variety of chemical features displayed in the minor groove is thought to be insufficient for specific recognition by a DNA-binding protein (Bewley *et al*., 1998). However, specific protein-DNA interactions via the minor groove have been described for other proteins as well, for example the human estrogen related receptor hERR2 and the integration host factor IHF of *Escherichia coli* (Gearhart *et al*., 2003; Wang *et al.*, 1995).
Another interesting aspect of ComK-binding, is the ability of the protein to alter the DNA-conformation in the region to which it binds. Intrinsic bending of the DNA is not required for specific recognition of the K-box by ComK. However, binding of a ComK-tetramer induces a significant bending of the DNA in the promoter area. This bending has been determined to be 65º for the *comF*-promoter and 75º for the *comG*-promoter (Hamoen *et al*., 1998).

Characteristics of the ComK protein
ComK is a small protein of 192 amino acids, with a calculated molecular weigth of 22.8 kDa and an isoelectric point of 7.7. ComK-homologues can be found in closely related *Bacillus* species, like *B. cereus, B. anthracis* and *B. licheniformis*, but also in bacteria like *Oceanobacillus iheyensis* and some *Listeria* species (fig. 6).

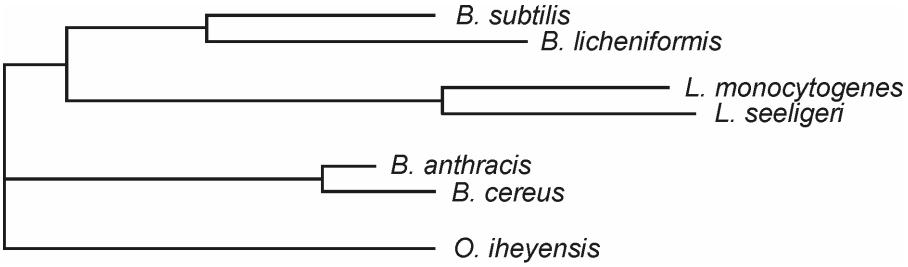


Figure 6. Schematic representation of homology between ComK-like proteins from different bacteria
All seven ComK-like proteins share clear homology throughout the entire amino acid sequence, with the exception of the N- and C-terminal regions. The C-terminal regions show homology between subgroups, which mainly determines the grouping as shown in this figure.

Furthermore, ComK shows limited sequence homology with the DNA-binding domain of hSRY, the human male sex determining factor. Interestingly, ComK and hSRY share some characteristics: i) both are regulatory proteins which bind to AT-rich sequences in the DNA, ii) protein-DNA interactions occur via the minor groove of the DNA and iii) both proteins induce bending of the DNA (Bewley *et al*., 1998; Geierstanger *et al*., 1994; Grosschedl *et al*., 1994; Hamoen *et al*., 1998). Despite the limited degree of homology with hSRY, the similarity in protein-DNA interactions and the use of various *in silico* analysis programs based on sequence comparisons, the location of the expected functional domains in ComK, *i. e.* the DNA-binding domain, the dimerization and tetramerization domains and the region interacting with MecA, have thus far not been identified.

Scope of this thesis

Extensive research over the past 15-20 years unraveled most of the complex regulatory pathway involved in the development of competence in *B. subtilis*. The transcriptional activator ComK was identified as the key regulatory protein in competence development and important determinants for regulation of and by ComK were elucidated, demonstrating that ComK activates gene transcription by binding to specific sites, K-boxes, upstream of ComK-regulated genes. The consensus sequence of these sites was de-

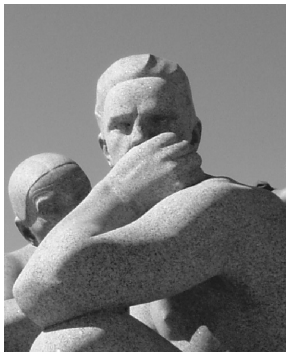
termined, showing a remarkable flexibility in the length of the spacing between the two repeats, the AT-boxes, of each K-box. Functional ComK was shown to be a tetramer, consisting of two dimers, which each bind to an AT-box. The variation in spacing length between both AT-boxes, requires a flexible binding nature of ComK to enable transcription activation at each class of ComK-regulated promoters.
In this thesis, the functionality of ComK in regulation of competence development was further investigated. **Chapter 2** describes the mechanism of transcription activation at the *comG*-promoter, demonstrating that the main role of ComK is to stabilize RNA-polymerase (RNAP)-binding to the promoter. Stabilization likely occurs via interactions of RNAP with upstream DNA, facilitated by ComK-induced bending of the promoter area. **Chapter 3** and **4** describe the investigation of ComK-DNA interactions. **Chapter 3** focuses on the search for functional domains in ComK, showing that the C-terminal region is required for transcription activation, by driving tetramerization, which is accompanied by DNA-bending. **Chapter 4** describes the importance of different elements in a K-box for ComK-binding and transcription activation. The most intriguing result consists of the fact that, despite the large natural variation in K-boxes of the *B. subtilis* genome, a single basepair change can dramatically lower the level of transcription activation by ComK, as was shown for the second thymine in an AT-box. **Chapter 5** describes the effects of the introduction of *B. subtilis* ComK into *Lactococcus lactis*, which contains a large number of K-boxes. DNA-microarray analyses show that ComK affects transcription of 7% of the *L. lactis* genome, mainly by disturbing normal transcription by binding to available K-boxes, which results in gene repression instead of gene activation, which normally occurs in *B. subtilis*. Finally, **chapter 6** discusses the main conclusions, suggestions for future work and the implications of the performed research.

Chapter 2

Mechanism of transcription activation by the competence transcription factor ComK of *Bacillus subtilis*

Kim A. Susanna, Aske F. van der Werff, Chris D. den Hengst, Belén Calles, Margarita Salas, Gerard Venema, Leendert W. Hamoen and Oscar P. Kuipers

This chapter is based on *Journal of Bacteriology*, **186,** 1120-1128



Abstract

The development of genetic competence in *Bacillus subtilis* is regulated by a complex signal transduction cascade, resulting in the synthesis of the competence transcription factor ComK. ComK is required for the transcription of the late competence genes, encoding the DNA-binding, -uptake and integration machinery. *In vivo* and *in vitro* experiments have shown that ComK is responsible for transcription activation at the *comG*-promoter. In this study, we investigated the mechanism of transcription activation. The intrinsic binding characteristics of RNA-polymerase (RNAP) with and without ComK at the *comG*-promoter were determined, showing that ComK stabilizes binding of RNAP. This stabilization probably occurs through interactions with the upstream DNA, since a deletion of this DNA resulted in an almost complete abolishment of stable RNAP-binding. Furthermore, the presence of an extra AT-box in addition to the common ComK-binding site was shown to be a strong requirement *in vitro*. *In vitro* transcriptions with *B. subtilis* RNAP reconstituted with wildtype α-subunits and with C-terminal truncation mutants of the α-subunits were performed, demonstrating that these truncations do not abolish transcription activation by ComK. This result indicates that ComK is not a type I activator. We also show that ComK is not required for open complex formation. A possible mechanism for transcription activation is proposed, implying that the major stimulatory effect of ComK is on binding of RNA-polymerase.

Introduction

Genetic or natural competence is a physiological differentiation state in which bacteria are able to take up exogenous DNA from the medium. This phenomenon has been extensively studied in the Gram-positive soil bacterium *Bacillus subtilis*. Competence development is initiated at the onset of the stationary growth-phase as a result of a complex regulatory cascade. Through quorum sensing, environmental signals, like nutrient availability and cell density, are sensed and interpreted (Dubnau, 1993; Hamoen *et al*., 2003b). Regulation by this cascade leads to the synthesis of the competence transcription factor ComK. Via an autoregulatory loop, ComK stimulates transcription of its own gene. In addition to this, ComK is required for the transcription of the late competence genes, *comC, E, F* and *G*, which encode the DNA-binding and -uptake machinery and *addAB* and *recA*, which are involved in DNA-recombination and -integration (Chen and Dubnau, 2003; Haijema *et al*., 1995 and 1996; Hoa *et al*., 2002; Mohan and Dubnau, 1990; Van Sinderen *et al*., 1995). ComK-dependent genes are characterized by the presence of a ComK-binding site, a K-box, in the promoter area. A K-box consists of two AT-boxes, which are separated by a spacing of a discrete number of helical turns. Previous research described the mechanism of ComK-binding to a K-box (Hamoen *et al*., 1998; Van Sinderen *et al*., 1995). A transcriptional fusion of the *comG*-promoter with *lacZ* showed that *in vivo* transcription of *comG* was completely abolished in a *comK*-deletion strain (Van Sinderen *et al*., 1994a). *In vitro* transcription studies confirmed that ComK alone is capable of activating transcription at the *comG*-promoter (Hamoen *et al*., 1998).
Transcriptional regulation by activators has been shown to affect transcription initiation at one or more of the following steps: i) stimulation of RNA-polymerase (RNAP)-binding, ii) stimulation of isomerization to an open promoter complex and iii) helping in promoter escape (Adhya and Garges, 1990; Adhya *et al*., 1993; Lloyd *et al*., 2001; McClure, 1985). The free energy of reaction intermediates of transcription initiation can be limiting at any of these steps. Activator interactions could function in lowering the energy barrier of the rate-limiting step(s) in order to accelerate the overall transcription initiation reaction (Geiselmann, 1997).
Possible interactions at a promoter site that could lead to transcriptional activation are generally divided into three separate le-

vels: A) direct protein-protein contacts between the activator and RNAP, B) conformational changes transmitted by the DNA upon binding of the activator and C) additional contacts with other DNA segments than the core promoter, such as interactions between the DNA sequence upstream of the activator-binding site and the backside of RNAP (Déthiollaz *et al*., 1996; Eichenberger *et al*., 1996). It has been postulated that the activation mechanism depends on the architecture of the promoter as well as on the rate-limiting step(s) in transcription initiation for that promoter (Hochschild and Dove, 1998).
In this study, the mechanism of ComK-dependent stimulation of transcription at the *comG*-promoter has been investigated. It was shown that RNAP-binding to the promoter is stimulated by ComK and that stabilization of RNAP-binding requires the presence of DNA upstream of the promoter region. Furthermore, transcription activation by ComK was not abolished when RNAP with C-terminal truncations of the α-subunits was used. Isomerization to an open complex promoter was shown to be ComK-independent. The implications of these results for the mechanism of transcription activation by ComK at the *comG*-promoter are discussed.

Materials and methods

Strains, plasmids and growth conditions
Bacterial strains and plasmids used in this study are listed in table 1. *B. subtilis* strains were grown in minimal medium (Venema *et al*., 1965), if required supplemented with 10 µg/µl kanamycin or 5 µg/µl neomycin. *E. coli* was grown in TY-medium (Sambrook *et al*., 1989), with 100 µg/µl ampicillin, if required. For overexpression of MBP-ComK, strains were grown in the presence of 0.2% glucose. Expression was induced with 0.3 mM IPTG.

Table 1. Bacterial strains and plasmids

Strain or plasmid	relevant features	Reference
Strains		
B. subtilis 8G5	*trpC2 his met tyr-1 ade nic ura*	Bron and Venema, 1972
B. subtilis NIG2001	*trpC2 pheA2 neo^r rpoC$_{his6}$*	Fujita and Sadaie, 1998
B. subtilis 8GG	*amyE::comG-lacZ* fusion	This work
B. subtilis 8GG6	*amyE*::*comG*+6-*lacZ* fusion	This work
E.coli BL21(DE3)	F$^-$ *ompT hsdS$_B$* (rB$^-$ m$_B^-$) *gal dcm*	Novagen
Plasmids		
pCD2	Overexpression of *B. subtilis* σA	Chang and Doi, 1990

Table 1, continued

Plasmid	relevant features	Reference
pAN-G	Ap^r, pAN583 with *comG*-promoter	Hamoen *et al*., 1998
pAN-G+6	Ap^r, mutant *comG*-promoter	This work
pAN-G-AT2-GC	Ap^r, random high GC DNA upstream *comG*. Deletion AT-box 3	This work
pAN-G-AT3-GC	Ap^r, random high GC DNA upstream of *comG*.	This work
pAN-G-AT2-sipS	Ap^r, *sipS* DNA upstream of *comG*. Deletion of AT-box 3	This work
pAN-G-AT3-sipS	Ap^r, *sipS* DNA upstream of *comG*	This work
pAN-G-AT2-codY	Ap^r, *L.lactis codY* DNA upstream of *comG*. Deletion of AT-box 3	This work
pAN-G-AT3-codY	Ap^r, *L.lactis codY* DNA upstream of *comG*	This work
pBTK-G	Ap^r, Km^r, *comG-lacZ* fusion	This work
pBTK-G+6	Ap^r, Km^r, *comG*+6-*lacZ* fusion	This work

DNA techniques, materials and transformations

Standard methods for molecular biology were used unless specified otherwise (Ausubel *et al*., 1998). Enzymes were purchased from Roche, New England Biolabs, Promega or Pharmacia. DNA oligonucleotides were synthesized by Gibco BRL or Invitrogen. Radiolabeled nucleotides were obtained from Amersham. Plasmid isolations and PCR-product purifications were performed using the High Pure Plasmid Isolation kit and the High Pure PCR Product Purification kit, respectively (both from Roche). Chromosomal DNA of *B. subtilis* was isolated as described before (Venema *et al*., 1965). *B. subtilis* was transformed as described by Anagnostopoulos and Spizizen (1961). *E. coli* strains were transformed using $CaCl_2$-induced competence (Sambrook *et al*., 1989).

PCR amplifications

PCR reactions were carried out as described by Innes and Gelfand (1990) using Pwo or Expand DNA-polymerase as enzyme (both from Roche). *B. subtilis* 8G5 chromosomal DNA, *Lactococcus lactis* chromosomal DNA or plasmid pAN-G were used as templates. Primers used in the PCR-reactions are listed in table 2. Probes for use in electrophoretic mobility shift assays were amplified by PCR. A combination of primers G1 and G2 was used to create the standard *comG*-promoter fragment. A combination of primers G2 and G1trn and of primers G2 and AT-AT3 resulted in a truncated *comG*-promoter fragment with two or three AT-boxes, respectively, but lacking the upstream DNA-region. Probes with longer upstream DNA sequences were made with primers G7 and G2 for wildtype *comG*-promoter DNA, retG-1 and G2 for high GC-DNA and SipS-XhoI or cod20 with G2 for high AT-DNA of *sips*- or *codY*-origin.

Table 2. Oligonucleotide primers used for PCR

Primer	Sequence
G1	5′-CCGGAATTCATGGTGACCATGTCTGCT
G2	5′-CGCGGATCCCTCTCCTTTCAACGC
G7	5′-TTTTGTGCAGCGTGCCCCGC
retG-1	5′-GATTTTTGTGATGCTCGTCAG
G+6F	5′-CTTTGTTTGATTACCTTTTCTTCTTTTTC
G+6R	5′-CTTGGGAAAACGTGATTTTGTGAGATG
G1trn-XhoI	5′-GATCCTCGAGAGAATTGGTTTTTCAGCATATAAC
G2trn-XhoI	5′-CTAGCTCGAGGGGTACCGAGCTCGAATTCG
G3trn-XhoI	5′-GATCCTCGAGGAAAGTCTTTTTTCTTGCCA
AT-AT2	5′-CATGGAAGACTGGGTAAGAATTGGTTTTTCAGCATATAAC
AT-AT3	5′-CATGGAAGACTGGGTAGAAAGTCTTTTTTCTTGCCAG
cod20	5′-ACACCATGGCTACATTACTTGAAAAAACACG
cod21	5′-ATAGAATTCCTCTGACTTTTAGAAATTACGTCG
SipS-XhoI	5′-CATGCTCGAGAACTGCCGGGAATATATTGG
SipS-BpiI	5′-CATGGAAGACTCTACCCACATCATGCC

Plasmid constructions

Plasmid pAN-G+6 was constructed using long-range PCR with primers G+6F and G+6R on pAN-G as template. Ligation of the subsequent PCR-product yielded pAN-G with a 6-bp insertion in between the K-box and the −35 sequence. Primers were designed to create a unique *Hin*dIII restrictionsite at the place of insertion. Wildtype and mutant *comG*-promoter fragments were cloned into the pBTK2 *amy*-locus integration vector (Meijer *et al*., 1995) as a *Bam*HI/*Eco*RI restriction fragment. The resulting plasmids pBTK-G and pBTK-G+6 were transformed into *B. subtilis* 8G5. β-galactosidase assays were performed as described (Van Sinderen *et al*., 1990).The pAN-G-AT-GC plasmids were constructed by PCR with pAN-G as template using primer G2trn-XhoI with G1trn-XhoI (for AT2) or G3trn-XhoI (for AT3). The products were digested with *Xho*I and self-ligated, resulting in plasmids in which the *comG*-upstream DNA is replaced by high GC-DNA (55% AT) of pUC-origin. Plasmids pAN-G-AT-*sipS* were made by PCR with primers G2trn-XhoI and AT-AT2 (for AT2) or AT-AT3 (for AT3). The products were digested with *Xho*I and *Bpi*I. Into this fragment, a PCR-product was ligated, made with primers SipS-XhoI and SipS-BpiI on *B. subtilis* chromosomal DNA as template and digested with *Xho*I and *Bpi*I. In the resulting plasmids, the *comG*-upstream DNA is replaced by *sipS*-upstream DNA (61% AT). Plasmids pAN-G-AT-codY were constructed by PCR on pAN-G as template with primers G2trn-XhoI and G1trn-XhoI or G3trn-XhoI for AT2 and AT3 respectively. Into this fragment, an internal gene fragment of *L. lactis codY* was ligated, amplified with primers cod20 and cod21. In the resulting plasmids, the *comG*-upstream DNA is replaced by DNA of *codY*-origin (60% AT). Use of primer G3trn or AT-AT3 resulted in the deletion of the third AT-box of *comG*.

Purification of σ^A-specific RNA-polymerase and of σ^A factor
To purify RNAP, an overnight culture of *B. subtilis* NIG2001 (Fujita and Sadaie, 1998) was diluted 100 times, grown at 37°C in 2xTY and harvested at T_0. All subsequent procedures were performed at 4°C. Cells were collected by centrifugation (10 min., 6000 rpm, Beckman centrifuge) and washed with ice-cold buffer A (20 mM Tris-HCl (pH 8.0), 0.2 M NaCl, 1 mM 2-mercaptoethanol, 10% glycerol, 10 mM $MgCl_2$). Cells extracts were obtained by French press in buffer A with 0.5 mM PMSF and centrifuged (20 min., 20,000 rpm, Beckman ultra-centrifuge, SW50 rotor), after adding an additional 0.5 mM PMSF. The supernatant was diluted 10 times in buffer A and applied to a Talon (clontech) or Ni-NTA (Qiagen) column. The loaded column was washed with buffer A and B (buffer A with 5 mM Imidazol) to remove non-specifically bound proteins. Bound proteins were eluted by increasing the imidazole concentration upto 50 mM in buffer A. Protein-containing fractions were diluted in low salt buffer (20 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 20% glycerol, 1 mM 2-mercaptoethanol) and applied to a prepacked, disposable 5 ml Heparin-agarose column (Pharmacia). After extensive washing with low salt buffer, RNAP was eluted by increasing the NaCl-concentration in buffer A upto 1.2 M. Finally the sample was dialysed against cold dialysis buffer (10 mM Tris-HCl (pH 8.0), 7.5% glycerol, 1 mM 2-mercapto-ethanol). Sigma A was purified from inclusion-bodies in *E. coli* as described (Chang and Doi, 1990). Before use in electrophoretic mobility shift assays or *in vitro* transcriptions, holo-enzyme was reconstituted on ice for at least 10 minutes by mixing RNAP and sigma A in a 1:1 molar ratio.

Electrophoretic mobility shift assays
For use in electrophoretic mobility shift assays (EMSA's), ComK was purified by the method of Hamoen *et al.* (1998). EMSA's were performed essentially as described (Van Sinderen *et al*., 1995). The PCR-generated DNA probes were endlabeled with T4-polynucleotide-kinase by use of [γ-^{32}P]-ATP. Purified proteins and probes were premixed on ice in 20 µl binding buffer (20 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 100 mM KCl, 0.5 mM dithiotreitol, 0.05 mg/ml poly[d(I-C)], 0.05 mg/ml BSA, 8.7% glycerol). All reactions were performed in the presence of 200 µM ATP and 200 µM GTP, with the exception of data in fig. 5, when indicated. Binary complexes were formed by incubating 15 minutes at 37°C. To distinguish open and closed RNAP-promoter complexes, 2 µl of a 0.3% heparin-solution was added directly prior to electrophoresis on a non-denaturing 4% polyacrylamide gel. Gels were run in TAE buffer (40 mM Tris-acetate (pH 8.2), 2 mM EDTA) at 100 V, dried, and autoradiographed.

In vitro *transcription assays*
In vitro transcription reactions contained in 25 µl, 25 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, 45 mM ammonium sulfate, 200 µM of each UTP, ATP and GTP, 80 µM [α-^{32}P]CTP (2 µCi), 1 µg of poly[d(I-C)], 1 µg of BSA and 9 nM template DNA. Templates used were supercoiled pAN-G (Hamoen *et al*., 1998) and derivatives and a 260-bp DNA fragment containing the phage Φ29 C2 promoter (Mencia *et al*., 1996), which give rise to transcripts of 360 and 98 nucleotides, respectively. RNAP's reconstituted with α-subunit truncation mutants (Δ15, Δ37, Δ59) were used as described by Mencia *et al.* (1996). ComK-protein was added to a final concentration of 0.35 µM. Reactions were performed at 37°C and processed as described (Monsalve *et al*., 1996; Nuez *et al*., 1992). Transcripts were separated by denaturing polyacrylamide gel electrophoresis and quantified using a Fuji BAS-IIIs image analyzer.

Results

ComK stimulates binding of RNAP at the comG*-promoter*
The basal prokaryotic promoter consists of four critical elements, *i. e*. the –35 and –10 hexamers, the spacer length between these two hexamers and upstream auxiliary elements.

Figure 1. Sequence of the *comG*-operon promoter region
Nucleotides are numbered relative to the transcription start site. Three potential AT-boxes are present (underlined). The boldface, underlined site is located at a position comparable with the K-box in other ComK-regulated promoters. The start of the common K-box (–86) and the extra AT-box (-109) are indicated. The –10 and –35 regions are printed boldface. The *B. subtilis* σ^A consensus sequence is shown underneath (capitals: in >70% of the promoters (Helmann, 1995)).

DNA-sequences resembling the consensus of such a core region are efficient binding sites for RNA-polymerase (RNAP). Nevertheless, they may be poor promoters in the absence of activator proteins (Bracco *et al*., 1989; Ellinger *et al*., 1994b; Frederick *et al*., 1995). The presence of a –35 consensus hexamer is important for efficient binding of sigma A-RNAP to the promoter, since sigma A (σ^A) makes specific contacts with the DNA at this region (Gross *et al*., 1992). In general, the homology score of promoter sequences correlates closely with the *in vitro* binding affinity of σ^A-RNAP (Brunner and Bujard, 1987; Ellinger *et al*., 1994b; Frederick *et al*., 1995). The transcription properties of σ^A-RNAP at the *comG*-promoter (fig. 1) were analyzed with *in vitro* transcription assays (Hamoen *et al*., 1998). It was shown that in the absence of ComK hardly any transcripts are formed, while the presence of ComK stimulates transcription upto 50-fold, indicating that ComK is sufficient and required to activate *comG*-transcription.

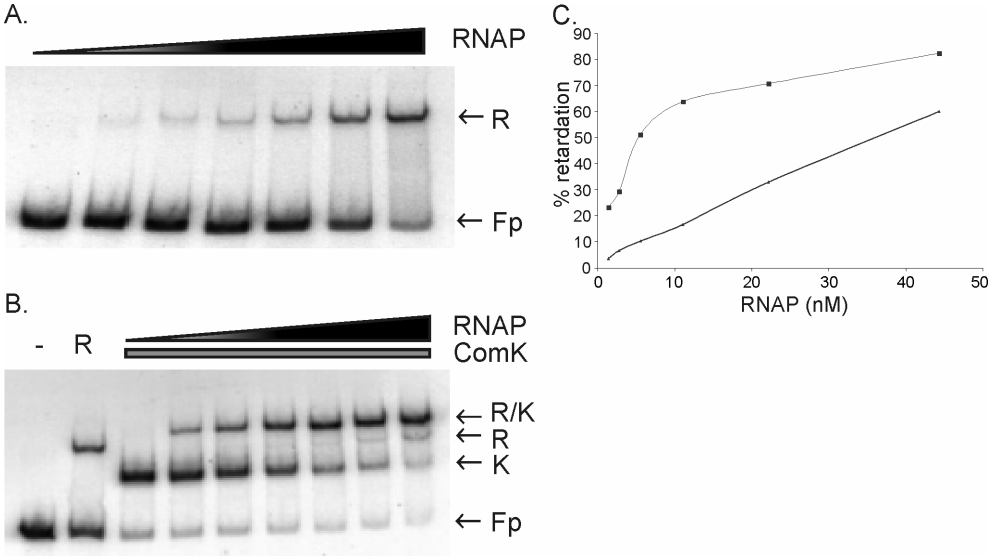


Figure 2. ComK stimulates RNAP-binding to the *comG*-promoter
EMSA's were performed with increasing RNAP concentrations (2-fold increments from 0 to 44 nM). (A) RNAP-binding to a 200-bp *comG*-promoter fragment in the absence of ComK. (B) RNAP-binding in the presence of 0.35 μM ComK. For comparison, a blanc sample (-) and only RNAP (R, 22 nM) are shown. K: ComK, R: RNAP, Fp: free probe. (C) Quantification of RNAP-binding. Percentage calculated by dividing the signal of the RNAP band by the total signal in each lane. Triangles: no ComK present; squares: 0.35 μM ComK.

Binding properties of σ^A-RNAP were analyzed with electrophoretic mobility shift assays (EMSA's). RNAP-binding can occur in the absence of ComK, but was increased 2 to 5 fold in the presence of ComK (fig. 2), resulting in a super-shifted complex. This suggests that ComK stimulates RNAP-binding to the *comG*-promoter.

RNAP-binding is stabilized by ComK via the upstream DNA
When a truncated *comG*-promoter fragment, lacking the DNA upstream of the common K-box, was used in an EMSA under the same conditions, stabilization of the RNAP-promoter complex was abolished (fig. 3A), although binding of ComK or RNAP alone was not disturbed. Therefore, it can be concluded that for stabilization of the complex, the DNA upstream of the K-box is important. Stimulatory effects of upstream DNA on the transcription activation process are known for several promoters (Eichenberger *et al*., 1996; Rao *et al*., 1994; Ross *et al*., 1993). Often, a specific activating sequence, the UP-element, can be distinguished, consisting of an AT-rich region located between -40 and -60 relative to the transcription start site (Estrem *et al*., 1998, 1999; Gourse *et al*., 2000). In the case of the *comG*-promoter, this region is occupied by ComK-binding, but the DNA-bending ability of ComK suggests a possible specific sequence to be located further upstream of the promoter (Pérez-Martin and De Lorenzo, 1997). To test whether the importance of the upstream DNA for stable RNAP-binding at the *comG*-promoter is a result of the presence of a specific sequence in this DNA or of the structural presence of the DNA, mutants were constructed in which the upstream DNA of *comG* was replaced by non-specific DNA, with either a high GC- or high AT-content. Furthermore, two types of constructs were tested, which differ in the number of ComK-dimer binding sites upstream of the promoter. Commonly, a K-box consisting of two AT-boxes is present upstream of ComK-activated genes (Hamoen *et al*., 1998). In the case of *comG*, an extra AT-box is located upstream of the common K-box, with an interval of two helical turns calculated from the start of the common K-box. In one set of mutants, called AT3, all three AT-boxes were included, while in the other set, AT2, only the two AT-boxes of the common K-box were present (fig. 3C). The different promoters were tested by EMSA's and *in vitro* transcription assays (fig. 3B), showing that deletion of the third AT-box resulted in an almost complete loss of stabilization of RNAP-binding and transcription in the presence of ComK.

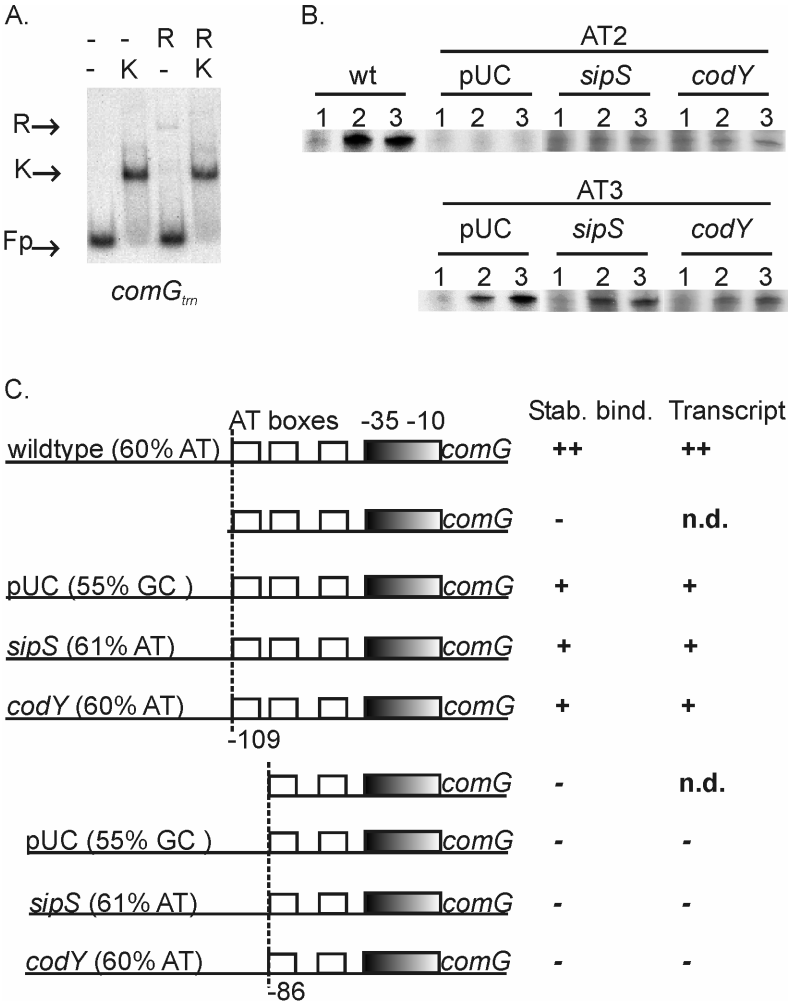


Figure 3. The upstream DNA region stabilizes the RNAP-promoter complex
(A) EMSA's were performed using a truncated *comG*-promoter ($comG_{trn}$, two AT-boxes, no upstream DNA). Reactions contained RNAP (17.5 nM) and/or ComK (0.35 μM) as indicated. K: ComK, R: RNAP, Fp: free probe. (B) *In vitro* transcription assays were performed as described in materials and methods on pAN-G or derivatives containing wildtype or non-specific upstream DNA, as indicated in panel C. AT2: two AT-boxes, AT3: three AT-boxes, 1: no ComK present, 2: 0.07 μM ComK, 3: 0.35 μM ComK. (C) Schematic overview of the constructs tested for the effect of the upstream DNA on stabilization of RNAP-binding and transcription. Squares: AT-boxes, rectangle: RNAP-binding site. The fragments contain upstream DNA (300 bp) of wildtype *comG*-origin, high GC- or high AT-DNA or no upstream DNA. The dotted lines indicate at which position of the wildtype

comG-promoter the fusions or truncations were made (numbers as in fig. 1). The situations with non-specific upstream DNA or no upstream DNA were tested for both three and two AT boxes. ++: wildtype stabilization of RNAP-binding/ transcription in the presence of ComK (100%), +: 40-80%, -: 0-10%.

This result indicates that the presence of the third AT-box is of great importance for (*in vitro*) transcription activation at the *comG*-promoter. However, the box alone is not sufficient to stabilize RNAP-binding, since stabilization is still almost completely abolished when a truncated *comG*-promoter, lacking the DNA upstream of the three AT-boxes is used. The replacement of *comG*-specific upstream DNA with either high GC- or high AT-DNA showed only a slight reduction of stabilization of RNAP-binding which never exceeded a 2 to 2.5 fold difference, demonstrating that the presence rather than a specific sequence of the upstream DNA is required for stabilization of RNAP-binding, followed by transcription activation at the *comG*-promoter.

ComK is not a type I transcriptional activator

An important class of prokaryotic transcription factors mediates transcription activation through direct contacts with the RNAP. A preferred activation target is the C-terminal domain of the α-subunit of RNAP (Boucher and Stibitz, 1995; Engelhorn and Geiselmann, 1998). In general, those activators binding at or upstream of position –60 relative to the transcription start site, normally interact with the α-subunit (Ishihama, 1993). To investigate whether ComK stimulates transcription through contacts with the α-subunit, *in vitro* transcription assays were performed using RNAP reconstituted with either wildtype α-subunit or α-subunits lacking the last 15, 37 or 59 amino acids from the carboxyl-end, respectively. Equivalent amounts of the reconstituted RNAP's were added to the transcription reactions and the products were separated by electrophoresis. The results demonstrated that the RNAP's containing truncation mutants of the α-subunit were still stimulated by ComK (fig. 4A), suggesting that direct protein-protein contacts between ComK and the α-subunits are not required for transcription activation. However, the maximum level of transcription by RNAP's reconstituted with mutant α-subunits was reduced approximately 2-fold compared to that of the wildtype polymerase (fig. 4B). This suggests that the C-terminal domain of the α-subunit is important for optimal transcription activity, as will be discussed.

Using EMSA's, it was also shown that ComK did not promote binding of purified α-subunits (wildtype and truncation mutants), nor of purified σ^A, to the promoter (results not shown). Therefore, it can be concluded that direct protein-protein contacts between ComK and the alpha or sigma subunits of RNAP are not required for stabilizing RNAP-binding to the promoter.

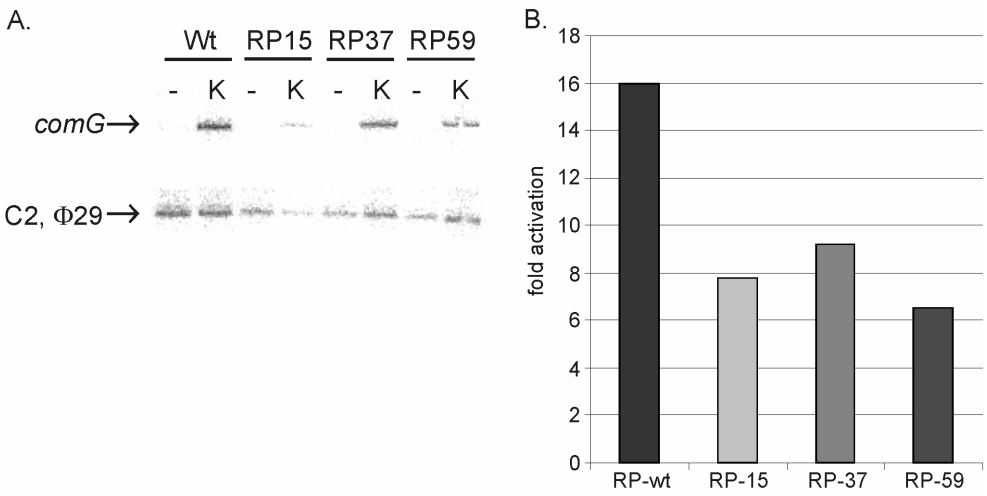


Figure 4. Stimulation of transcription by RNAP with wt or mutant α-subunits
(A) *In vitro* transcription reactions were performed as described in materials and methods. Products were separated by electrophoresis. The constitutive C2 promoter from phage Φ29 was used as an internal standard for specific activity of the reconstituted RNAP's. (B) Transcription products were quantified with a Fuji BAS-IIIs image analyzer. Transcription is reported as fold stimulation over that with no ComK added for each RNAP preparation, corrected for differences in specific RNAP activity by using the internal standard. RP-wt, RP-15, RP-37, RP-59: RNAP's reconstituted with wildtype α-subunit and Δ15, Δ37, Δ59 C-terminal truncation mutants, respectively.

Open complex formation is independent of ComK

The second step in transcription initiation is the formation of an open complex. Competitor resistance is widely used as a functional assay for the open complex formation (Straney and Crother, 1985; Whipple and Sonenshein, 1992). Challenge experiments using heparin indicated that the presence of ComK is not required for open complex formation at the *comG*-promoter (fig. 5A). The formation of open complexes was shown to be dependent on the presence of the initiating nucleotides. Only when ATP and GTP

were added to the reaction mixtures, RNAP-promoter complexes became resistant to a heparin challenge. Upon addition of both nucleotides, an additional stabilizing effect on RNAP-binding and an additional shift were observed, compared with the complexes formed in the absence of nucleotides or in the presence of only ATP or GTP.

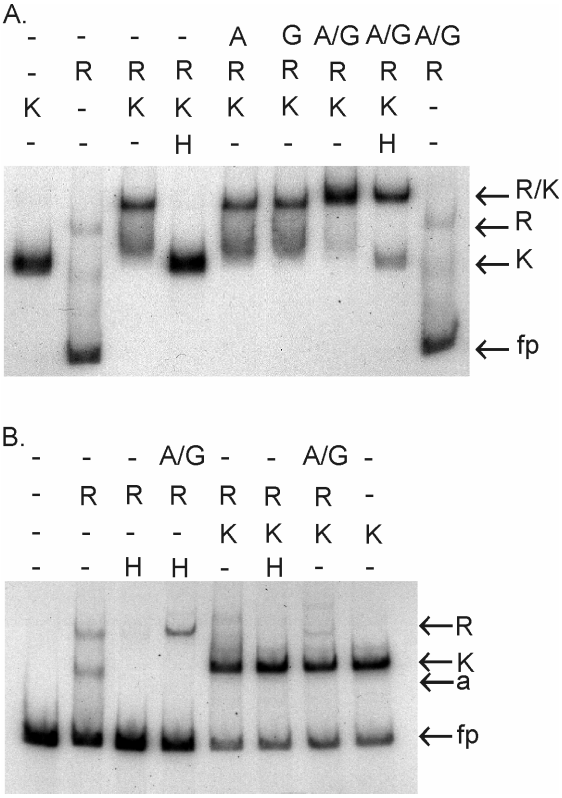


Figure 5. Open complex formation and helix face dependency
(A) Requirement of initiating nucleotides for heparin resistance. EMSA's were performed using the ^{32}P-labeled *comG*-promoter fragment. (B) Helix face dependency of the ComK effect on RNAP-binding to the *comG*-promoter. EMSA's were performed using the ^{32}P-labeled *comG*+6 (206-bp) promoter fragment. (A) and (B) Reactions contained mixtures of RNAP (17.5 nM), ComK (0.35 μM) and initiating nucleotides (200 μM) as indicated above the lanes. Heparin challenges were performed as described in materials and methods. The positions of the different complexes are shown on the right. K: ComK, R: RNAP, H: heparin, A: ATP, G: GTP, fp: free probe. a: aspecific binding in lane 2, connected with the purified RNAP sample.

The additional shift could be caused by the fact that in the presence of both initiating nucleotides a short abortive transcript can be formed (Straney *et al*., 1989). Likely, the formation of a short transcript stabilizes binding of RNAP and causes a slightly altered migration pattern upon electrophoresis. Normal open complex formation is also seen when the truncated *comG*-promoter with two AT-boxes was used in heparin challenge experiments (results not shown). This confirms that, when RNAP is bound to the promoter, isomerization to an open promoter complex occurs upon the addition of nucleotides and independent of ComK.

Transcription activation is helix face dependent

To investigate whether the orientation of a K-box is important for transcription activation, a *comG*-promoter fragment was created with a 6-bp insertion in between the K-box and the –35 hexamer. In this situation, the K-box is located on the opposite face of the DNA-helix, and as a consequence, bound ComK is located on the opposite face of the helix compared to the downstream RNAP. It has been found that ComK induces a bend in the DNA upon binding (Hamoen *et al*., 1998). In the mutant *comG*+6 construct, the induced bend is present in the opposite direction compared to wildtype. The *comG*+6 promoter fragment was used in *in vitro* transcription assays, showing abolishment of transcription activation by ComK. *In vivo*, this promoter was placed in front of the *lacZ*-gene in the *amy*-locus of the *B. subtilis* genome. β-galactosidase expression was abolished to the same level as in a *comK*-deletion strain (results not shown; Van Sinderen *et al*., 1994a). EMSA's with *comG*+6 showed that stabilization of RNAP-binding was disturbed, a situation comparable with the truncated *comG*-promoters. Still, the initial level of RNAP-binding in the absence of ComK was the same as in the wildtype *comG*-promoter, as was open complex formation (fig. 5B). Binding of RNAP seemed to be lost, when ComK binds to the other side of the helix. These results indicate a strict helix-face-dependency for transcription activation by ComK. It has been proposed that intrinsic or protein-induced DNA-bending immediately upstream of a promoter can activate transcription by looping the upstream DNA sequences around to interact with the backside of RNAP (Parekh and Hatfield, 1996; Pérez-Martin and Espinosa, 1991; Pérez-Martin and De Lorenzo, 1997). This would explain both the requirement for the upstream DNA of *comG* and the helix-face-dependency.

Discussion

Transcription initiation frequently requires the interaction of several DNA-binding proteins that ultimately modulate the activity of RNAP. In competence development in *B. subtilis*, *comK* encodes the central regulator, the competence transcription factor ComK. ComK activates and binds specifically to the promoters of the late competence genes and the genes required for recombination. *In vitro* studies have shown that purified ComK alone is capable of activating transcription at the *comG*-promoter. In this report we describe its mechanism of transcription activation.
In order to determine in which step of transcription initiation ComK is involved, several approaches were taken. Using EMSA's, it was shown that RNAP can bind to the *comG*-promoter also in the absence of ComK, but that the amount of closed complexes is stimulated upto 5-fold when ComK is present (fig. 2). Stabilization of RNAP-promoter complexes in the presence of ComK was shown to depend on the DNA upstream of the K-box. When this DNA was deleted, binding of ComK and RNAP alone were not disturbed (fig. 3A), but the super-shifted RNAP-promoter complex was no longer stabilized. This suggests that stabilization of RNAP-binding is a result of bending of the upstream DNA by ComK, thereby enabling interactions between the DNA and RNAP.
Replacement of the *comG*-specific upstream DNA by either high GC- or high AT-DNA resulted in only a slight reduction in stabilization of RNAP-binding and transcription (fig. 3C). Several sequence comparisons were made between the upstream DNA of *comG* and other ComK-activated genes. No clear conserved sequences could be indicated, but an important difference between *comG* and other ComK-activated genes is the presence of an extra AT-box upstream of the common K-box of *comG*. Binding studies and *in vitro* transcriptions comparing promoter fragments with two or three AT-boxes upstream of *comG*, demonstrated the requirement of the third box for stabilization of RNAP-binding and transcription *in vitro* (fig. 3C). Previous footprinting studies by Hamoen *et al.* (1998) demonstrated that all three AT-boxes are protected by ComK. The presence of this extra AT-box might be the determinant that results in the large transcription activity at the *comG*-promoter. Array studies indicated that *comG*-transcription is the highest of all ComK-activated genes and *in vitro* transcription studies with ComK-activated genes have thus far only been success-

ful for *comG* (Berka *et al*., 2002; Hamoen *et al*., 2002; Ogura *et al*., 2002). Studies with a truncated *comG*-promoter fragment that still contained all three AT-boxes, but lacked the upstream DNA, no longer showed stabilization of RNAP-binding, indicating that in addition to the third AT-box, also the presence of more upstream DNA is required.
The requirement for the upstream DNA, correlates with the results shown in *in vitro* transcription assays with RNAP's reconstituted with wildtype or C-terminal truncation mutants of the α-subunit. The results indicated that a direct interaction between the α-CTD and ComK is not required for RNAP activation (fig. 4), since ComK could still stimulate transcription by mutant RNAP's. EMSA's showed that ComK is not able to recruit purified α-subunit or σ^A to the *comG*-promoter, another indication that no significant contacts between ComK and RNAP are involved in transcription activation. Although activation was not abolished in the reconstituted mutant RNAP's, a twofold reduction of maximal transcription was observed. Rowe-Magnus *et al.* (1998) reported a similar observation for transcription of the *spoIIG*-promoter by SpoOA~P. They suggested an effect on the interaction of RNAP with promoter DNA by the α-subunit mutation. The C-terminal domain of the α-subunit is known to interact with additional promoter sequences (UP-elements) to stabilize polymerase-DNA interactions at some promoters (Eichenberger *et al*., 1996; Estrem *et al*., 1998, 1999; Gourse *et al*., 2000; Rao *et al*., 1994; Ross *et al*., 1993). Although a clear UP-element could not be demonstrated in the upstream region of the *comG*-promoter, it is still possible that specific AT-rich stretches in the upstream DNA interact with RNAP. In *E. coli* the same residues of the α-C-terminal domain were found to be involved in interactions with activators such as CRP and promoter UP-elements (Murakami *et al*., 1996). If the α-subunit C-terminal domain would indeed help to stabilize binding of RNAP to the promoter, it would explain why truncations of this domain disturb optimal transcription activity and why the presence of upstream DNA is important for optimal RNAP-binding. Although we cannot totally rule out the possibility that ComK interacts with RNAP through some other region of the enzyme than the α-subunits or σ^A, like the β- or β'-subunits, we currently favour the notion that ComK activation of transcription from the *comG*-promoter is mediated via stabilization of RNAP-binding by the upstream region of the promoter DNA.

Each step in the transcription initiation process can in principle be a target for regulation by transcriptional activators (Geiselmann, 1997). Transcription activation can involve multiple interactions between a single activator molecule and the transcription machinery, each interaction being responsible for a specific mechanistic consequence. In fact, such multiple interactions have become a commonly observed feature in transcription activation (Busby and Ebright, 1997; Hochschild and Dove, 1998; Niu *et al*., 1996). To advance our understanding of the effect of ComK on transcription initiation at the *comG*-promoter, several experiments were performed to investigate if ComK is involved in other steps as well, in addition to stimulating binding of RNAP. When the orientation of the ComK- and RNAP-binding sites was inverted in the *comG*+6 promoter construct, stabilization of RNAP-binding to the promoter was abolished. In the presence of ComK, no basal level of RNAP-binding was observed, suggesting that binding of ComK to the opposite face of the helix hinders RNAP-binding to the promoter. In addition, transcription from this promoter was lost, both *in vitro* and *in vivo*. Helix-face-dependency has been taken as evidence for crosstalk between RNAP and the activator protein (Giladi *et al*., 1992). Since protein-protein interactions with the α or σ subunits of RNAP are not involved in transcription activation by ComK, it is likely that the orientation of the DNA-bend induced by ComK-binding is responsible for the helix-face-dependency. It has been proposed that activator-induced bending of DNA upstream of the promoter facilitates caging of RNAP to optimize the promoter (Adhya *et al*., 1993; Buc, 1986; Pérez-Martin and De Lorenzo, 1997). We conclude that the mechanism of activation relies on contacts between the DNA upstream of the ComK-binding sites and the backside of RNAP. Similar findings have been reported for the *gal*- and *lac*-promoters (Déthiollaz *et al*., 1996) and for the CRP-dependent *malT*-promoter (Eichenberger *et al*., 1997).
For *in vitro* transcription assays supercoiled templates were used, since these were found to be approximately 20-fold more efficient in transcription than run-off transcription assays using linear templates (results not shown). The supercoiled state of the chromosome is known to affect the activity of many promoters (Parekh *et al*., 1996). It is a fairly common phenomenon among bacterial promoters to be stimulated by DNA-superhelicity (Borowiec and Gralla, 1987; Pruss and Drlica, 1989). The stimulatory effect of superhelicity of the template on transcription efficiency is also in

agreement with our model. The influence of DNA-bending on regulatory processes may be modulated by superhelicity (Gartenberg and Crothers, 1991). Specifically, supercoiling and bending of DNA may synergistically enhance polymerase contacts by creating a defined DNA-topology at the promoter site, a view also put forward by Zinkel and Crothers (1991). Alternatively, DNA-supercoiling may optimize the three-dimensional geometry of the DNA for correct alignment of the proteins and/or DNA sites, thus lowering energy barriers in transcription initiation (Lewis *et al*., 1999).
The *comG*-promoter has a strong resemblance to the *B. subtilis* σ^A-dependent consensus promoter (fig. 1). In general, such consensus-like promoters stably bind RNAP and require alterations to accelerate the late steps of the transcription initiation pathway (Ellinger *et al*., 1994a; Hsu, 2002). Therefore, ComK might also affect transcription initiation in one of the later stages after closed complex formation. The stabilization of the closed complex by ComK will, of course, contribute to accelerate the overall transcription process. The second step in transcription initiation, open complex formation, was found to be ComK-independent. Addition of initiating nucleotides was demonstrated to be sufficient to induce a heparin-resistant promoter complex. The isomerization to an open promoter complex is not disturbed when half a helical turn is inserted in between the promoter and the K-box. All of this clearly indicates that open complex formation is not a rate-limiting step for transcription initiation at the *comG*-promoter.
It has been suggested that consensus σ^A-promoters that efficiently bind RNAP and that exhibit a strong open complex formation, may be limited in the subsequent movement of the polymerase to the elongating complex (Carpousis and Gralla, 1985; Hsu, 2002). RNAP-binding at these promoters generates a nucleoprotein complex that is too stable to allow promoter escape (Hsu, 2002; Monsalve *et al*., 1996). Melting of the DNA-strands in the promoter area in the presence of nucleotides leads to an initiating complex that is trapped in short abortive transcripts synthesis (Menendez *et al*., 1987). Escape from this complex into an elongating transcription machinery involves major conformational changes, including loss of the promoter specific contacts and release of the sigma factor (Knaus and Bujard, 1988; Mishra and Chatterji, 1993; Straney and Crothers, 1985). Escape from abortive initiation has been demonstrated to be rate-limiting at several other prokaryotic promoters (Menendez *et al*., 1987; Smith and Sauer, 1996). In

the case of the *comG*-promoter, initial experiments were performed to elucidate the role of ComK in promoter escape of RNAP. To distinguish between an effect of ComK on RNAP-binding or on promoter escape, ComK had to be added after the binding step. In this case, involvement of ComK in promoter clearance could not be demonstrated, since transcription levels were severely decreased when ComK was added in a later stage of transcription initiation than in the binding step. Further research remains required to investigate whether bending of the upstream DNA by ComK results not only in stabilization of RNAP-binding, but also in creating optimal conditions for the following steps in the transcription initiation process, like promoter escape.
The proposed model for the role of ComK in transcription activation is summarized in fig. 6. Although RNAP is capable to bind to the *comG*-promoter in the absence of ComK, binding is stimulated when ComK is present (step i). In the case of the *comG*-promoter, ComK can bind to three AT-boxes, resulting in bending of the upstream DNA around ComK. This DNA probably interacts with RNAP, thus stabilizing the RNAP-promoter complex (step ii). Further studies should be performed to see whether interactions between the upstream DNA and the back-side of RNAP also help to induce conformational changes in the promoter DNA and/or RNAP that are required for promoter escape.

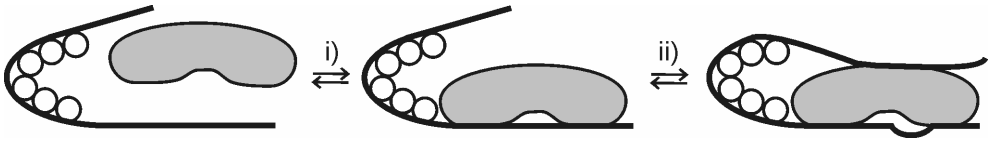


Figure 6. Model of transcription activation by ComK at the *comG*-promoter
ComK-dimers are represented as small circles, RNAP as a large ellipsoid and the DNA by a solid line. For details see text.

In this study, the mechanism of transcription activation by ComK was investigated at the *comG*-promoter. This promoter differs from the promoters of most other ComK-activated genes by containing a third AT-box. However, the transcription activation mechanism at promoters containing only two AT-boxes, is expected to be comparable with the model presented in this study. Likely, the major effect of ComK is stabilization of RNAP binding via the upstream DNA region.

Acknowledgements

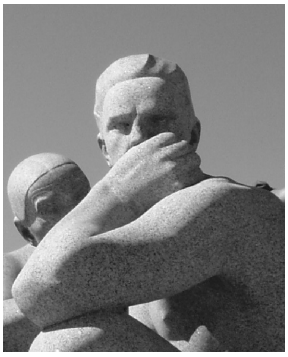
We would like to thank M. Fujita and Y. Sadaie for their kind gift of the strain *Bacillus subtilis* NIG2001, used for purification of RNA-polymerase and B. Chang and R. Doi for their kind gift of plasmid pCD2, used for overexpression and purification of sigma A. We also thank Sierd Bron, Caroline Eschevins and Wiep Klaas Smits for helpful discussions.

Chapter 3

The C-terminal region of the competence transcription factor ComK of *Bacillus subtilis* is required for transcription activation

Kim A. Susanna, Fabrizia Fusetti, Andy-Mark W. H. Thunnissen, Leendert W. Hamoen and Oscar P. Kuipers

This chapter has been submitted for publication



Abstract

The competence transcription factor ComK is the master regulator of competence development in *Bacillus subtilis*. In the regulatory pathway controlling competence development, ComK is involved in different interactions: i) protein-DNA interactions to stimulate transcription of ComK-dependent genes and ii) protein-protein interactions, which can be divided into interactions with other proteins, like MecA, for the formation of the proteolytic complex with ClpCP and interactions between ComK-proteins involving oligomerization into dimers and tetramers. The fact that ComK displays different types of interactions, suggests the presence of specific, distinct domains in the protein.
This chapter describes a search for functional domains in ComK, performed by constructing ComK-truncation variants, which were tested for DNA-binding, oligomerization and transcription activation. Truncations at the C-terminal end of ComK demonstrated the requirement of this part for transcription activation, but not for DNA-binding. Likely, the C-terminal region of ComK is involved in oligomerization of ComK-dimers into tetramers, as can be concluded from the fact that a C-terminal 25 amino acids truncation mutant displays only dimer binding. Surprisingly, a ComK-truncation variant lacking 9 amino acids from the N-terminal end showed higher transcription activation than wildtype ComK, when overexpressed in *L. lactis*. However, in *B. subtilis*, transcription activation by ComKΔN9 is two-fold lower than by wildtype ComK, resulting from a 5 to 6-fold lower protein level of ComKΔN9. Thus relatively ComKΔN9 is more active in transcription activation than wildtype ComK. These results suggest that the presence of this N-terminal extension on ComK is a trade-off between high transcription activation and a, thus far unidentified, role in regulation of ComK.

Introduction

The competence transcription factor ComK is the key regulatory protein in competence development in *Bacillus subtilis*. Genetic competence is a differentiation process, initiated at the onset of stationary growth, which enables the cell to take up and incorporate exogenous DNA in order to acquire new genetic abilities. Development of competence is tightly regulated via a complex regulatory system, centred around ComK (reviews: Dubnau and Lovett, 2002; Hamoen *et al*., 2003b). During exponential growth, competence development is prevented by controlling the level of ComK in the cell via both transcriptional and post-translational control. Transcription of *comK* is repressed by binding of AbrB, CodY and Rok to the promoter of *comK* (Hamoen *et al*., 2003a; Hoa *et al*., 2002; Serror and Sonenshein, 1996). Any ComK that is synthesized at this stage, is bound by MecA, an adaptor protein that targets ComK for proteolytic degradation by the ClpCP-protease complex (Turgay *et al*., 1998). At the onset of stationary growth, the cell responds to environmental changes, like nutrient deprivation and increased cell density, by relieving transcriptional repression of *comK* by AbrB and CodY (Hahn *et al*., 1995; Serror and Sonenshein, 1996) and by synthesis of ComS. This small protein binds to MecA, thereby replacing ComK, which results in the release of ComK from the proteolytic complex (D'Souza *et al*., 1994; Hamoen *et al*., 1995; Solomon *et al*., 1995). Once ComK is free in the cell, it activates transcription of its cognate gene, by binding to the promoter region, where it can overcome repression by Rok (Hoa *et al*., 2002; Van Sinderen and Venema, 1994b; Van Sinderen *et al*., 1995). Via this autostimulatory loop, intracellular ComK-levels increase rapidly and, subsequently, ComK activates transcription of other genes, *e. g.* the late competence genes, encoding the DNA-binding, -uptake and -integration machinery (Van Sinderen and Venema, 1994b).
ComK activates transcription by binding to specific sequences, K-boxes, located upstream of ComK-dependent genes. Each K-box consists of two AT-boxes, separated by a spacing of two, three or four helical turns between the starting basepairs of the repeating AT-box units. Functional ComK consists of a tetramer, composed of two dimers, which both bind to an AT-box. Binding of ComK is accompanied by DNA-bending (Hamoen *et al*., 1998). The major role of ComK in transcription activation is on stabilization of RNA-

polymerase binding, probably by facilitating interactions with the upstream DNA-region, through bending of the promoter region (chapter 2).
As becomes clear from the regulatory pathway, ComK is involved in different types of interactions: i) protein-DNA binding, which is of particular interest because the interactions occur through the minor groove of the DNA-helix (Hamoen *et al*., 1998), and ii) protein-protein interactions, involving either interactions with other proteins, like MecA, or interactions between ComK-proteins resulting in dimerization and tetramerization. The occurrence of these different interactions suggests that the ComK-protein contains specific, distinct domains, responsible for one or more of the interactions. ComK is a relatively small protein of only 22.8 kDa. Amino acid sequence comparisons and structural prediction studies did not reveal the presence of any clear domains. However, an alignment with ComK-like proteins from other bacterial species shows a large degree of conservation throughout the entire amino acid sequence, with the exception of the extreme N- terminal part (fig. 1). The C-terminal region is not conserved throughout all listed bacteria, but shows some conservation when subgroups are concerned, for example between the *Listeria* species or between *B. subtilis* and *B. licheniformis*.
In this study we focus on the role of specific parts of the N- and C-terminal regions of ComK in oligomerization, DNA-binding and transcription activation. For this purpose, His-tagged ComK, wild-type and N- or C-terminal truncation variants, were expressed in *Lactococcus lactis*, in which transcription activation was investigated using a (*B. subtilis*) *comG-lacZ* reporter fusion. DNA-binding of the proteins was demonstrated using electrophoretic mobility shift assays. This approach was chosen to make a clear dissection between the effect of the truncations on DNA-binding, oligomerization and transcription activation. Furthermore by using an inducible system in *L. lactis*, effects of MecA-binding and interactions with other *B*. *subtilis* regulators could be circumvented.
Using this approach, we demonstrate that the C-terminal region of ComK is required for transcription activation, but not for DNA-binding. Most likely, a loss of tetramerization occurs in the C-terminal truncation variants. Interestingly, truncation of 9 amino acids from the N-terminal end of ComK results in a ComK-variant that is more active in transcription activation in *L. lactis*. To determine the biological relevance of this region, the ComKΔN9 variant

was introduced in *B. subtilis*, where it showed a relatively higher transcription activation ability, despite a lower protein expression level. Protein levels could be increased by introducing a *mecA*-deletion, although the expression level remained lower than expected. The fact that ComKΔN9 displays higher transcription activation, but lower protein expression, suggests that the presence of this N-terminal extension on ComK is a trade-off between high transcription activation and a not yet elucidated regulatory role involved in ComK stability or activity.

```
B.subtilis    MSQKTDAPLESYEVNGATIAVLPEE-IDGKICSKIIEKDCVFYVNMKPLQIVDRSCRF  57
B.licheniform MS-TEDMTKDTYEVNSSTMAVLPLG-EGEKPASKILETDRTFRVNMKPFQIIERSCRY  56
B.anthracis   MENKVERYVENYVVNKNTMALLPVILSEKKIVTRVVEVQDSFFVFQKPLDIIERSCRK  58
B.cereus      MESKVERYVENYVVTKNTMALLPVILSEKKIVTRVVEMNDSFFVFQKPLDIIERSCRK  58
O.iheyensis   --MKKQQYLE---ITPFTLAVVSEQDENGRYIAKVLEEEAEYVLEAKPTNVIDYACKY  53
L.monocytogen -MKKEQISTQFYEVNPHTMIIFPKK-SGSIVYSEIYEVDSHYTSKFTPFELIKTSCNF  56
L.seeligeri   -MKKELISNQFYEVNPHTMIIFPKK-SGSIVYSEIYEVDSQYTSKFTPFELIKTSCNF  56
                 .     :   :.  *: :..         :.: * :  :    .* :::. :*.

B.subtilis    FGSSYAGRKAGTYEVTKISHKPPIMVDPSNQIFLFPTLSSTRPQCGWISHVHVKEFKA 115
B.licheniform FGSSYAGRKAGTYEVIKVSHKPPIMVDHSNNIFLFPTFSSTRPQCGWLSHAHVHEFCA 114
B.anthracis   HGSSFLGRKEGTKELTHITHKAPIAISPTDQLYFFPTYSYSRKECAWLSHFYIESNKE 116
B.cereus      HGSSFLGRKEGTKELTHITHKAPIAISPADQLYFFPTYSYSRKECAWLSHFYIESNKE 116
O.iheyensis   FGASLKGRQEGTREISGITHKAPITIDPASGMYFFPTKSPSNANCSWIAHSHIKEVRR 111
L.monocytogen FGSSYEGRKEGTKHLIGVTHKPPIIIDPVTSTYVFPTVAPSSTECIWIFPQHIKDYHA 114
L.seeligeri   FGSSYDGRKEGTKHLIGVTHKPPIIVDPVTSTYVFPTAAPSPTDCIWIFPQHIKDYQT 114
              .*:*  **: ** .:  ::**.** :.     :.*** : :  :* *:   ::..

B.subtilis    TEFDDTEVTFSNGKTMELPISYNSFENQVYRTAWLRTKFQDRID-HRVPKRQEFMLYP 172
B.licheniform AKYDNTFVTFVNGETLELPVSISSFENQVYRTAWLTTKFIDRIEGNPMQKKQEFMLYP 172
B.anthracis   LKDGNLIIRFINGFAVKLEISKTSFENQQNRTAKLRTEYEDRRKKQGNPCFKEVDKNE 174
B.cereus      LKDGNLIIRFINGFAVKLEISKTSFENQQNRTAKLRTEYEDRRKKQGNPCFKEVDKKE 174
O.iheyensis   EEEGKTSVLFKNGSKISLDISHGSLINQLQRTAQIRFLLDDRIKRWRKPPHSEDDKDS 169
L.monocytogen IGFNHTLITFSNMETFEIDMSLASFNNQIARTSMLHMKFSQKMRMMESNFPSMNRFFP 172
L.seeligeri   IGYNHTLIRFSNQQTFEIDMSLASFNNQIARTSMLHMKFSQKMRMMESNFPSLNMFFP 172
                 ..  : * *   ..: :* *:  **  **: :     ::         .

B.subtilis    KEERTKMIYDFILRELGERY---------- 192
B.licheniform KEDRNQLIYEFILRELKKRY---------- 192
B.anthracis   ESRLKPAYESVYFVKEEEV---------- 193
B.cereus      ESTLRPAYESVYFVKEGEV---------- 193
O.iheyensis   ---------------------------- 169
L.monocytogen PTTLAAEPRRYYSTMLPNNEEPNDPQDPEQ 202
L.seeligeri   PATIAAETRRYYNTMILEND---DPTDSDQ 199
```

Figure 1. Overview of homology between ComK-proteins from different bacteria Amino acids are conserved throughout the entire protein, except for the N- and C-terminal parts. The C-terminal region shows conservation between subgroups (grey-scale shadings). *, : and . indicate identical residues in all sequences, conserved substitutions and semi-conserved substitutions, respectively. Alignments were made using ClustalW (www.ebi.ac.uk/clustalw).

Materials and Methods

Bacterial strains, media and growth conditions
Bacterial strains used in this study are listed in table 1. *E. coli* strains were grown in TY-medium, supplemented with ampicillin (100 μg/ml), when required. For purification of the MBP-ComK fusion protein, cells were grown in the presence of 0.2% glucose and 0.3 mM IPTG to induce overexpression (pMal protein-fusion and purification system, NEB). *L. lactis* strains were grown at 30ºC in 2-fold diluted M17-based medium (Difco), supplemented with 0.5% glucose (GM17) and appropriate antibiotics (chloramphenicol and/or erythromycin, both 4 μg/ml). Protein overexpression was achieved by inducing the nisin-inducible promoter with 1:10,000 dilutions of supernatant of an overnightculture of nisin-producing *L. lactis* NZ9700 in GM17 (De Ruyter *et al*., 1996).

Table 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant features	Reference
Strains		
E. coli XL1blue	*endA1 gyrA96 thi hsdR17 (rK-mK-) supE44 relA1 lac/F' proAB laclq lacZΔM15 Tn10*	Stratagene
L. lactis NZ9000	MG1363 *pepN::nisRK*	Kuipers *et al*., 1998
L. lactis NZ9700	Nisin producing, containing Tn5276	Kuipers *et al*., 1993
B. subtilis 168	*trpC2*	Anagnostopoulos and Spizizen, 1961
B. subtilis BV2004	Δ*comK*, Sp^r	Hamoen *et al*., 2002
B. subtilis QB4650	Δ*mecA*, Km^r	Msadek *et al*., 1998
B. subtilis 168 a*myE::comKΔN9*	*amyE::comKΔN9*; Cm^r	This work
B. subtilis pG9	*amyE::comKΔN9*; pG-wt, Cm^r, Ery^r	This work
B. subtilis pG9ΔK	pG9 + Δ*comK*, Cm^r, Ery^r, Sp^r	This work
B. subtilis pG9ΔK-ΔM	pG9ΔK + Δ*mecA*, Cm^r, Ery^r, Sp^r, Km^r	This work
B. subtilis pGwt	pG-wt, Ery^r	Chapter 4
B. subtilis pGwt-ΔM	pG-wt, Δ*mecA*, Ery^r, Km^r	This work
Plasmids		
pMal-c2X	Amp^r, IPTG inducible *malE* with MCS	New England Biolabs
pMal-ComK	Amp^r, IPTG inducible *malE-comK*	This work
pDL	Cm^r, plasmid for integration in *amyE*	Yuan, Wong (1995)
pDL-ComKΔN9	Cm^r, *comKΔN9* region flanked by *amyE*	This work
pNZ8048	Cm^r, inducible expr. vector with P_{nisA}	Kuipers *et al*., 1998
pNZ-His_6-ComK	Cm^r, *his_6-comK* under control of P_{nis}	This work
pNZ-ComK-His_6	Cm^r, *comK-his_6* under control of P_{nis}	This work
pNZ-ComKΔC25	Cm^r, *his_6-comKΔC25* under P_{nisA} control	This work
pNZ-ComKΔC35	Cm^r, *his_6-comKΔC35* under P_{nisA} control	This work
pNZ-ComKΔC50	Cm^r, *his_6-comKΔC50* under P_{nisA} control	This work
pNZ-ComKΔN9	Cm^r, *comKΔN9-his_6* under P_{nisA} control	This work
pILORI4	Ery^r, pIL252 with MCS and promoter-less *lacZ* of pORI13	Larsen *et al.,* 2004
pG-wt	Ery^r, *comG-lacZ* with common K-box	Chapter 4

DNA manipulations, materials and transformations
Standard molecular biology methods were used as described (Ausubel *et al*., 1998; Sambrook *et al*., 1989). Enzymes were obtained from Roche, New England Biolabs or Pharmacia. Radiolabeled nucleotides were purchased from Amersham. For isolation of plasmids and purification of PCR products, the high pure plasmid isolation kit and the high pure PCR product purification kit, respectively, were used (Roche). Chromosomal DNA of *B. subtilis* was isolated as described by Venema *et al*. (1965). *E. coli* strains were transformed using $CaCl_2$-induced competence (Sambrook *et al*., 1989). Electrotransformations of *L. lactis* cells were performed using a gene-pulser (BioRad Laboratories) as described by Leenhouts and Venema (1993). *B. subtilis* strains were transformed as described by Anagnostopoulos and Spizizen (1961).

PCR amplifications and plasmid constructions
PCR reactions were performed as described (Innes and Gelfand, 1990), using Pwo or Expand DNA-polymerase (both from Roche) on chromosomal DNA of *B. subtilis* 168 as a template. Plasmids and primers used in this study are listed in table 1 and table 2, respectively. For purification purposes, an MBP-ComK fusion was constructed in *E. coli*. To construct this fusion, the *comK*-gene was amplified with primers ComK-start and ComK-end. The resulting PCR product was digested with *XmnI* and *Hind*III and cloned into *Xmn*I/*Hind*III digested pMal-c2X, yielding plasmid pMal-ComK.

Table 2. Oligonucleotide primers used for PCR
Restriction sites are printed boldface. The annealing sequences are underlined.

Primer	Sequence
comK-start	5'- ATGAGTCAGAAAACAGACGC
comK-end	5'- GGGG**AAGCTT**CGGACAAGGCACGCCGCC
H6Xa-ComK	5'- CATG**CCATGG**GGCACCATCACCATCACCATATCGAGGGAAGGATGA GTCAGAAAACAGACGC
comK-Nice1	5'- GCAT**TCATGA**GTCAGAAAACAGACGC
comK-H6	5'- GATC**AAGCTT**CTAATGGTGATGGTGATGGTGGCCATACCGTTCCCC GAGCTCAC
comK-dN9-Nice	5'- GATC**CCATGG**AATCGTATGAAGTGAACGGCGC
stop1	5'- CCGC**AAGCTT**ACTATTCCTGTCTTTTCGGCACGCG
comK-dC35	5'- CCGC**AAGCTT**CTATCTGTCTTGGAATTTGGTTCTGAG
comK-dC50	5'- CCGC**AAGCTT**CTAGTTCTCGAACGAATTATAAGAGATCGGC
yhxC-H3	5'- GCCG**AAGCTT**ATCAGGACGCCGAGGAA
comK-start-BbsI	5'- GATC**GAAGAC**ATCATATTATGGCCTCCATCCTTTTTCTGC
comK-dN9-BbsI	5'- GATC**GAAGAC**ATTATGGAATCGTATGAAGTGAACGGCGC
comK+500	5'- GATC**GGATCC**ACTCGGCGGACTTGATGTGC
comG-AT2-EcoRI	5'- GATC**GAATTC**AGAATTGGTTTTTCAGCATATAACATCTCAC
comG-end-XbaI	5'- GATC**TCTAGA**TTATGCCTCTTCAATCAAGTTTTTGC

His-tagged variants of wildtype and mutant ComK were constructed in *L. lactis*. To yield wildtype ComK (wtComK) with an N-terminal His-tag, the *comK*-gene was amplified with primers H6Xa-ComK and ComK-end and digested with *Nco*I and *Hind*III. The PCR-product was cloned into *Nco*I/*Hind*III digested plasmid pNZ8048, resulting in pNZ-His_6-ComK. For wtComK with a C-terminal His-tag, primers ComK-Nice1 and ComK-H6 were used to amplify *comK*. After digestion with *Rca*I and *Hind*III, the PCR product was cloned into *Nco*I/*Hind*III digested pNZ8048, yielding plasmid pNZ-ComK-His_6. ComK truncation variants were constructed by amplifying the *comK*-gene with primers ComK-dN9-Nice and ComK-H6 for an N-terminal 9 amino acids truncation and primers H6Xa-ComK combined with Stop1 or ComK-dC35 for the truncation of 25 or 35 amino acids from the C-terminus, respectively. The PCR-products were digested with *Nco*I/*Hind*III and cloned into pNZ8048, also digested with *Nco*I/*Hind*III, yielding pNZ-ComKΔN9, pNZ-ComKΔC25 and pNZ-ComK ΔC35, respectively.
To produce an N-terminal 9 amino acids truncation variant of ComK in *B. subtilis*, a *comKΔN9* copy was integrated in the *amyE*-locus of *B. subtilis* 168 under control of the *comK*-promoter. For this purpose, the *comK*-region was amplified with primers yhxC-H3 and ComK-start-*Bbs*I and with ComK-dN9-*Bbs*I and ComK+500. Both PCR-products were digested with *Bbs*I and ligated to form a 2.3 kb product. Using the ligation-mixture as a template, the *comKΔN9*-region was amplified with primers yhxC-H3 and ComK+500. The PCR-product was digested with *Bam*HI and *Hind*III and ligated into *Bam*HI/*Hind*III digested pDL, resulting in the integrative plasmid pDL-ComKΔN9. Transformation of this plasmid to *B. subtilis* 168 resulted in strain *B. subtilis amyE*::*comKΔN9*. To test transcription activation by ComKΔN9, the strain was transformed with the replicative plasmid pG-wt, yielding strain *B. subtilis* pG9. Finally to have ComKΔN9 as the only ComK-protein in the cell, the wt*comK*-copy was deleted by transformation with chromosomal DNA of *B. subtilis* BV2004 (Hamoen *et al*., 2002), resulting in strain *B. subtilis* pG9ΔK. The control strain for transcription activation by wtComK consisted of *B. subtilis* 168 transformed with pG-wt, yielding strain *B. subtilis* pGwt.
To investigate potential differences in interactions of wtComK and ComKΔN9 with the proteolytic system, a *mecA* deletion was introduced in *B. subtilis* pG9ΔK and *B. subtilis* pGwt via transformation with chromosomal DNA of *B. subtilis* QB4650 (Msadek *et al*., 1998). This transformation resulted in strains *B. subtilis* pG9ΔK-ΔM and *B. subtilis* pGwt-ΔM, which contain ComKΔN9 or wtComK as the only ComK-protein in the cell, respectively.

Overexpression and purification of MBP-ComK
Overexpression of the MBP-ComK fusion protein was induced by adding IPTG (final concentration 0.3 mM) to an exponentially growing *E. coli* culture (OD_{600} of 0.7 - 0.8). After two hours of induction, cells were harvested by centrifugation (10 min., 8000 rpm, Beckman centrifuge). ComK was purified as described (Hamoen *et al*., 1998), yielding either native ComK (after cleavage with factor Xa) or the MBP-ComK fusion protein.

Overexpression and cell extracts of ComK in L. lactis
Overexpression of wtComK and mutants with an N- or C-terminal His-tag was induced from the nisin-inducible promoter by adding supernatant from an overnightculture of *L. lactis* NZ9700 in 1:10,000 dilutions to *L. lactis* cultures at the end of the exponential growth. To obtain ComK for electrophoretic mobility shift assays, cells were harvested after two hours of induction (10 min., 8000 rpm, Beckman centrifuge). Cell extracts were prepared by resuspending the cell pellet from 2 ml of culture in 200 µl resuspension buffer (10 mM EDTA, 50 mM Tris-HCl, pH 8). Cells were disrupted by fast prep (45 sec, speed 6, Bio101 analyzer).

Purification of His_6-ComK overexpressed in L. lactis
Overexpression of His_6-ComK was induced in *L. lactis* at the end of the exponential growth-phase as indicated above. Purification of His_6-ComK was performed under denaturing conditions. Cells from 1 litre of culture were resuspended in 6 ml buffer A (0.1 M NaH_2PO_4, 10 mM Tris-HCl, pH 8.0, 2% Triton X-100) with lysozyme (1 µg/µl). After 20 minutes incubation at 37°C, cells were disrupted by French Press. A clear cell extract was obtained by ultracentrifugation (25 min., 25.000 rpm, 4°C, Beckman ultracentrifuge, SW-41 rotor). Purification of His_6-ComK was performed in a batch procedure, mixing 2.5 ml talon column material (clontech) with the cell extract, to which buffer B (Buffer A + 8 M Urea) was added to a final volume of 50 ml. After 1 hour incubation at room temperature, the sample was centrifuged (10 min., 10,000 rpm, Eppendorf centrifuge). The column material was washed subsequently with buffer B and C (8 M Urea, 0.1 M NaH_2PO_4, 10 mM Tris-HCl, pH 6.3). His_6-ComK was eluted in 1 ml buffer E (8 M Urea, NaH_2PO_4, 10 mM Tris-HCl, pH 4.5) and dialysed against dialysis buffer with gradually decreasing Urea concentrations. When reaching 1 M Urea, the sample was dialysed overnight at 4°C against dialysis buffer without Urea (20 mM Tris-HCl, pH 8, 1 mM EDTA, 0.5 mM DTT, 0.1 M Na_2SO_4). The purified protein was divided into aliquots and stored at -80°C.

Transcription activation test system in L. lactis
Transcription activation by wtComK and ComK-variants was first tested in *L. lactis*. ComK-expression was induced at the end of the exponential growth-phase from the nisin-inducible promoter on the pNZ-plasmids (De Ruyter *et al*., 1996). Transcription activation was tested using the *lacZ* reporter under control of the *B. subtilis comG*-promoter on plasmid pG-wt (chapter 4). Samples were taken at the moment of induction until two hours later, with time-intervals of 30 minutes. β-galactosidase activity was determined as described (Israelsen *et al.*, 1995). Control samples were analysed for protein expression with SDS-PAGE (Laemmli *et al*., 1970) and Western blotting (Towbin *et al*., 1979). His-tagged ComK-proteins were detected with a His_6-specific first antibody and an anti-mouse horseradish peroxidase secondary antibody (both from Amersham) and visualized by chemiluminescent detection using the ECL Western blotting analysis system (Amersham).

Transcription activation test system in B. subtilis
Transcription activation by ComKΔN9 was compared with wtComK in *B. subtilis* using the *lacZ*-gene on plasmid pG-wt as reporter. Samples for β-galactosidase assays were taken from transition point until 4 hours in the stationary growth-phase, with one-hour intervals. Expression of wtComK and ComKΔN9 were checked with Western blot analysis, using a ComK-specific first antibody (Van Sinderen and Venema, 1994b) and an anti-rabbit horseradish peroxidase secondary antibody (Amersham).

Electrophoretic mobility shift assays
Electrophoretic mobility shift assays (EMSA's) were performed essentially as described (Hamoen *et al*., 1998). A *comG*-promoter fragment was used as a probe, amplified by PCR with primers comG-AT2-*Eco*RI and comG-end-*Xba*I. The product was end-labeled with T4 polynucleotide kinase using [γ-^{32}P]-ATP. EMSA's were performed with cell extracts and compared with shifts from pure ComK, as indicated in the figures. Proteins and DNA-probes were premixed on ice in 20 µl binding buffer (20 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 100 mM KCl, 0.5 mM dithiotreitol, 0.05 mg/ml poly[d(I-C)], 0.05 mg/ml BSA and 8.7% glycerol). Protein-DNA complexes were allowed to form during 15 minutes incubation at 37°C, followed by running 15 µl of each sample on a non-denaturing 6% polyacrylamide gel. Gels were run in TAE buffer (40 mM Tris-acetate (pH 8.2) and 2 mM EDTA) at 100V, dried, and autoradiographed using a phosphor-screen (Packard). Shifts were quantified with the software package of Quantity one.

Results

B. subtilis ComK is active in transcription in L. lactis

In order to determine the characteristics of the ComK truncation mutants, a test system was developed to discriminate between effects on transcription activation, DNA-binding and oligomerization. This was achieved by using an *in vitro* system to investigate DNA-binding and oligomerization (discussed below) and an *in vivo* system in *L. lactis* to determine transcription activation. For this purpose, wildtype and mutant *B. subtilis* ComK were produced in *L. lactis* under control of the nisin-inducible promoter. As reporter for transcription activation, a (*B. subtilis*) *comG*-promoter-*lacZ* fusion on plasmid pG-wt was used. Furthermore, by using *L. lactis*, possible effects on ComK of other *B. subtilis* regulators, like MecA, could be circumvented, which enables a clean investigation of the transcription activation abilities of the ComK-mutants.

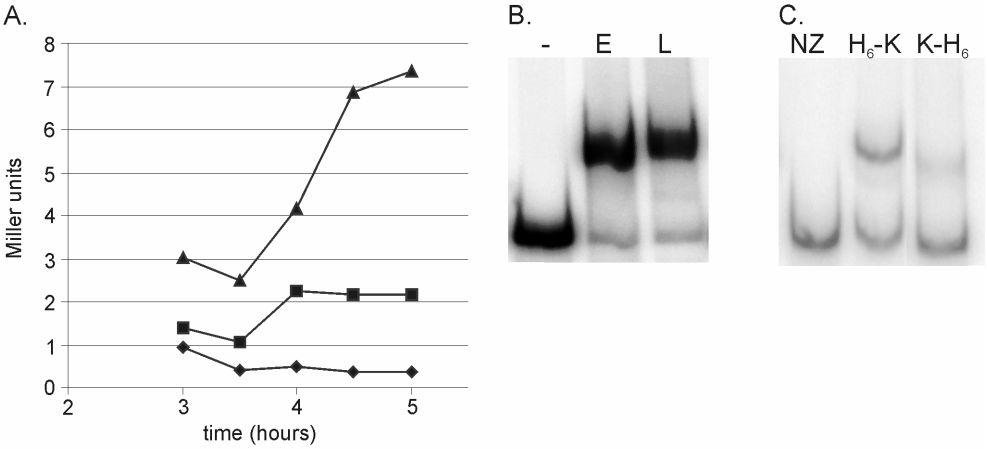


Figure 2. ComK of *B. subtilis* is active when expressed in *L. lactis*
(A). β-galactosidase assays were used to monitor transcription activation by His_6-ComK and ComK-His_6. ComK-expression was induced in *L. lactis* at t=3 by adding nisin. A *comG*-promoter-*lacZ* fusion was used as a reporter for regulation by ComK. Transcription is observed in cells expressing His_6-ComK (triangles) or ComK-His_6 (squares), but not in cells with empty plasmid pNZ8048 (diamonds). (B). DNA-binding of purified His_6-ComK was shown using EMSA's. The *comG*-promoter (with common K-box) was used as a probe. Binding of His_6-ComK purified from *L. lactis* (L) was compared with ComK purified from the MBP-fusion system in *E. coli* (E). Both proteins show a similar shift relative to unbound DNA (-). (C). Binding of His-tagged ComK was tested directly from extracts of induced cells with pNZ8048 (NZ), pNZ-His_6-ComK (H_6-K) or pNZ-ComK-His_6 (K-H_6).

Before the transcription test system could be used for the truncation variants, it was established for wildtype ComK (wtComK). For this purpose, wtComK was expressed with a His-tag fusion to the N- (His_6-ComK) or C-terminus (ComK-His_6). β-galactosidase studies demonstrated that transcription activation at the *comG*-promoter was stimulated by both His-tag ComK-variants, indicating that ComK of *B. subtilis* is active in *L. lactis* (fig. 2A). However, transcription activation by His_6-ComK is 3 to 4 times higher than by ComK-His_6, despite less difference (at most 2-fold) in protein expression (results not shown), suggesting that the position of the His-tag influences the ability of ComK to activate transcription at a K-box. Furthermore, transcription is about 10-fold lower than normally observed for wtComK without a tag in *B. subtilis*.

ComK isolated from L. lactis *shows normal binding*

Extensive research on ComK revealed the binding characteristics of purified ComK from *B. subtilis* using electrophoretic mobility shift assays (EMSA's) (Hamoen *et al*., 1998). In the present study, ComK was overexpressed as a His-tag fusion, which might influence DNA-binding or oligomerization of the protein. In order to determine the effect of the His-tag, His_6-ComK was isolated from *L. lactis* using a talon column. However, all His_6-ComK ended up in the unbound fraction, indicating that the His-tag is inaccessible for binding to the column. To solve this problem, purification was performed under denaturing conditions, using buffers with 8 M Urea. Under these conditions, His_6-ComK did bind to the column and could be purified. The purified protein could successfully be renatured to active His_6-ComK, as was demonstrated by EMSA's (fig. 2B). The binding characteristics of His_6-ComK were comparable to binding of ComK purified from the MBP-ComK system in *E. coli*, as was described before by Hamoen *et al*. (1998). Since His_6-ComK could only be purified under denaturing conditions, an alternative system was tested, using cell extracts. Although the observed band is weaker than for purified proteins, DNA-binding with typical ComK-characteristics could be demonstrated using cell extracts of *L. lactis* overproducing either His_6-ComK or ComK-His_6 (fig. 2C). This possibility provided a quick tool to test binding of ComK truncation variants directly from cell extracts. Furthermore, the use of this system prevents wrong interpretations of binding characteristics of mutants, due to potential unsuccessful renaturing of the truncation variants.

C-terminal region of ComK is required for transcription activation
As can be seen in fig. 1, ComK-proteins from different bacterial species show a clear sequence conservation throughout the entire protein, with the exception of the N-terminal end. The C-terminal part is not conserved for all ComK-proteins, but shows conservation when smaller subgroups are concerned, *e. g. B. subtilis* and *B. licheniformis*, of which the majority of the last 35 amino acids is identical or similar. Since the C-terminal region is partially conserved, a specific role for this part of ComK is expected. To investigate this role, truncations lacking 25 or 35 amino acids from the C-terminus were constructed and overexpressed as N-terminal His-tag fusions in *L. lactis*. β-galactosidase assays were used to determine transcription at the *comG*-promoter, demonstrating that transcription activation by both His_6-ComKΔC25 and ΔC35 is completely abolished (fig. 3A), despite a normal production level as shown by SDS-PAGE and Western blot experiments (fig. 3C).

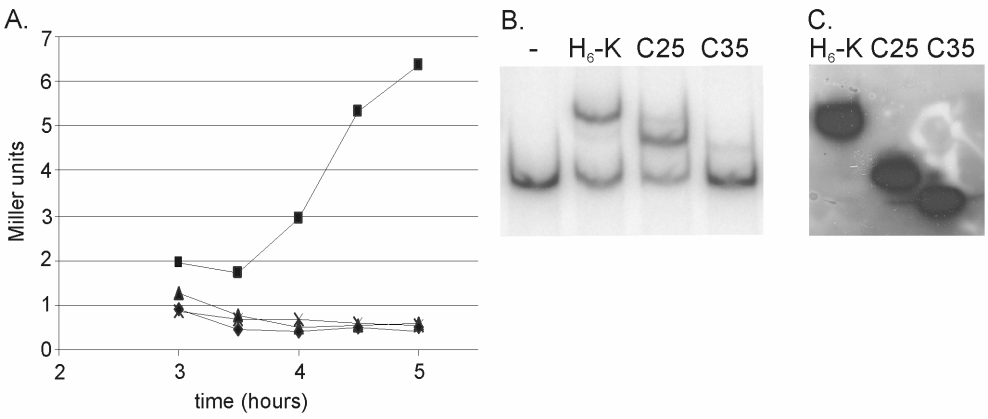


Figure 3. The C-terminal region of ComK is required for transcription activation (A). β-galactosidase assays were performed, comparing His_6-ComK (squares) with His_6-ComKΔC25 (crosses) and ΔC35 (triangles) in *L. lactis*. As a negative control *L. lactis* pNZ8048 (diamonds) was used. Protein overexpression was induced at t=3. β-galactosidase activity in cells overexpressing truncation variants is comparable with the negative control, demonstrating the requirement of the C-terminal part of ComK for transcription activation. (B). EMSA's were performed with cell extracts demonstrating binding of wtComK and the C-terminal truncation variants, although with a smaller shift for His_6-ComKΔC25 and ΔC35. (C). Western blots were performed using a His_6-specific first antibody to compare expression levels of the different ComK variants in *L. lactis*. (B): -: blanc. (B) and (C): H_6-K; His_6-ComK; C25: His_6-ComKΔC25; C35: His_6-ComKΔC35.

The loss of transcription activation by the ComK-mutants might be due to a loss of DNA-binding or to a defect in the transcription activation mechanism. To discriminate between these possibilities, EMSA's were performed with cell extracts containing His_6-ComK, His_6-ComKΔC25 or His_6-ComKΔC35, demonstrating that both ComK-mutants could still bind to the DNA of the *comG*-promoter (fig. 3B). However, the observed shift by binding of His_6-ComKΔC25 and His_6-ComKΔC35 is smaller than the control shift of His_6-ComK and, for His_6-ComKΔC35, the band is much weaker.

Effect of C-terminal truncations on tetramerization of ComK
Since the loss of transcription activaton by the C-terminal truncation variants of ComK is not due to a loss of DNA-binding, it is expected that the C-terminal region of ComK is involved in another aspect of ComK-interactions. An interesting hypothesis would be that the C-terminus of ComK is required for tetramerization of the protein, which would explain the smaller shift observed for the mutants as binding of a single ComK-dimer due to the loss of the cooperative binding of two dimers forming a tetramer. To be able to separate tetramerization from dimer binding, EMSA's were performed using a probe of the wildtype *addAB*-promoter and one in which five basepairs were introduced in between both AT-boxes (addAB+5). Introduction of five basepairs, corresponding to half a helical turn, positions both AT-boxes (and thus both bound ComK-dimers) on opposite sides of the DNA-helix, thereby eliminating the possibility to form a tetramer (Hamoen *et al*., 1998). Using this approach, it was demonstrated that small differences in protein size, due to the truncations, only resulted in small differences in shift when the AT-boxes were located on opposite sides of the DNA-helix (fig. 4, lower panel). This can by no means explain the large differences in shifts, observed when both AT-boxes are located on the same side of the DNA-helix (fig. 4, upper panel). Furthermore, the shifts by His_6-ComKΔC25 and ΔC35 are similar on both probes, indicating that in both situations, these ComK-mutants bind as dimers instead of tetramers. The wtComK-shift on the other hand increases from the *addAB+5* to the *addABwt* K-box, corresponding with oligomerization from dimers into tetramers. For wtComK-binding, tetramerization is accompanied by bending of the DNA. A loss of tetramerization for the C-terminal truncation variants and thus a decrease of induced bending would explain the loss of transcription activation, since previous studies

demonstrated that the main effect of ComK in transcription activation at the *comG*-promoter is on stabilization of binding of RNA-polymerase, by facilitating interactions with the upstream DNA through bending of the promoter region (chapter 2).

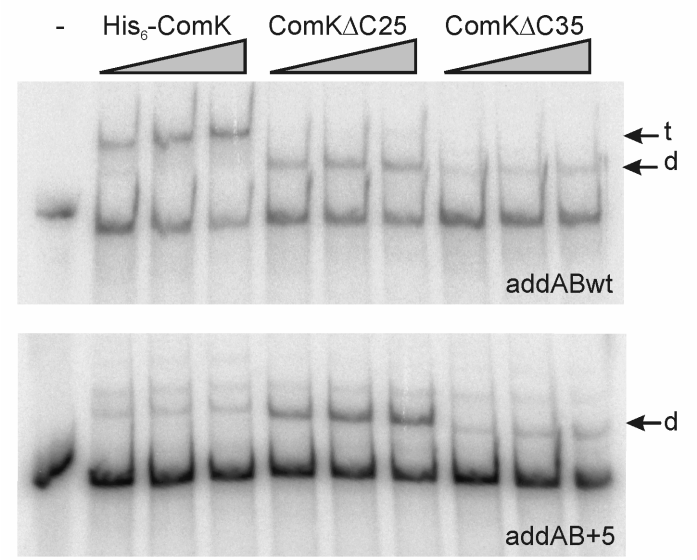


Figure 4. ComK C-terminal truncations bind as dimers instead of tetramers
Binding of His_6-ComK, His_6-ComKΔC25 and His_6-ComKΔC35 was compared on a K-box with both AT-boxes on the same side of the DNA-helix (addABwt) and a K-box with the boxes on opposite sides of the helix (addAB+5). Increasing cell extract amounts were used (0.5; 1 and 2 µl) for each protein overexpression, indicated with the concentration bar. -: no cell extract added. Arrows indicate the positions of ComK-dimer (d) and ComK-tetramer (t) binding. Both gels were run under the same conditions and for equal times.

Hyperactive ComK by truncation of 9 aa from N-terminus
As was shown in fig. 1, the N-terminal part of ComK is not strictly conserved throughout all ComK-proteins of the various bacterial species. This observation raises curiosity about the function of this protein region. As a first attempt to elucidate the role of the N-terminal part of ComK in different interactions, a truncation of 9 amino acids from the N-terminus was constructed and overexpressed in *L. lactis* (His-tag fusion to C-terminus). The ability of the mutant, ComKΔN9-His_6, to activate transcription was tested as described for the C-terminal truncations. Surprisingly, transcription activation by ComKΔN9-His_6 was increased about 2-3 fold

when compared to wtComK-His_6 (fig. 5A). As for the other ComK-mutants, DNA-binding by ComKΔN9-His_6 was determined using EMSA's, demonstrating that this is also increased compared with ComK-His_6 (fig. 5B), while Western blots demonstrated similar expression levels for both proteins (results not shown).

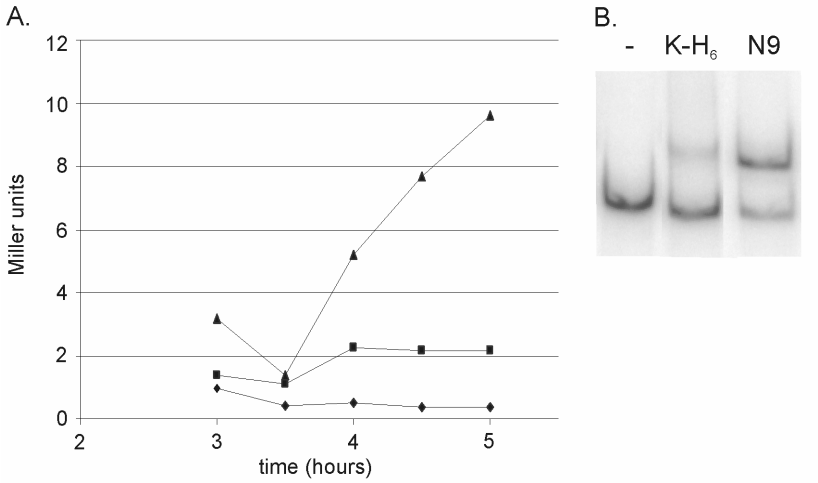


Figure 5. An N-terminal truncation results in a hyperactive ComK protein
(A). β-galactosidase assays were performed to determine transcription activation at the *comG* promoter-*lacZ* fusion. ComK-overexpression was induced in *L. lactis* at t=3. As a negative control, *L. lactis* pNZ8048 (diamonds) was used. Induction of ComKΔN9-His_6 (triangles) results in a 2-3 fold higher transcription activation than overexpression of ComK-His_6 (squares). (B). EMSA's were performed to investigate DNA-binding to the K-box of the *comG*-promoter. DNA-binding from an extract of cells overexpressing ComKΔN9-His_6 was compared with *L. lactis* overexpressing ComK-His_6. -: blanc; K-H_6: ComK-His_6; N9: ComKΔN9-His_6.

ComKΔN9 is relatively more active than wtComK in B. subtilis
The transcription test system in *L. lactis* provides a quick method to investigate the characteristics of the ComK truncation mutants. However, compared with *B. subtilis*, the test system in *L. lactis* is incomplete, considering all known other regulatory inputs, like for example the autostimulatory loop involved in *comK*-expression and the proteolytic degradation system consisting of MecA/ClpCP. The finding of a hyperactive ComK-variant raises the question whether this characteristic is specific for the truncation of the extreme N-terminal region or whether it results from a lack of regulatory input in the test system in *L. lactis*. In order to answer this question, ComKΔN9 was expressed in *B. subtilis*. The constructed

test strain, *B. subtilis* pG9ΔK, contained the *comK*-region with ComKΔN9 under control of the wt*comK*-promoter integrated into the *amyE*-locus, plasmid pG-wt to determine transcription activation and a deletion of the wildtype copy of *comK*. As a control *B. subtilis* pGwt was constructed, containing wtComK and plasmid pG-wt. β-galactosidase assays with the two strains demonstrated that the level of transcription activation by ComKΔN9 is about 2-fold lower than for wtComK (fig. 6 and fig. 7). However, Western blot analysis showed that the expression level for ComKΔN9 was about 5 to 6-fold lower than for wtComK (fig. 7B), indicating that, also in *B. subtilis*, ComKΔN9 is relatively more active in transcription activation than wtComK.

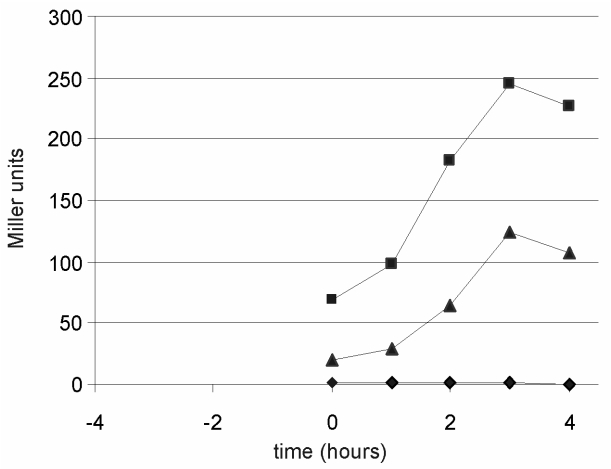


Figure 6. Transcription activation by ComKΔN9 is lower than by wtComK
β-galactosidase assays were performed to determine transcription activation by wtComK and ComKΔN9 in *B.* subtilis, using a *comG*-promoter-*lacZ* fusion on plasmid pG-wt as a reporter. Both wtComK and ComKΔN9 were transcribed under control of the *comK*-promoter, located upstream of their respective genes. As a negative control, *B. subtilis ΔcomK* was used. Squares: *B. subtilis* pGwt; triangles: *B. subtilis* pG9ΔK; diamonds: *B. subtilis* Δ*comK*.

Lower protein level is likely not due to altered MecA-interactions
As was described above, expression of ComKΔN9 in *B. subtilis* is lower than wtComK-expression. A possible explanation could be that the truncation of 9 amino acids from the N-terminus alters the interaction between ComK and MecA, which targets ComK for proteolysis. To investigate this possibility, a *mecA*-deletion was introduced in *B. subtilis* pG9ΔK, resulting in strain pG9ΔK-ΔM.

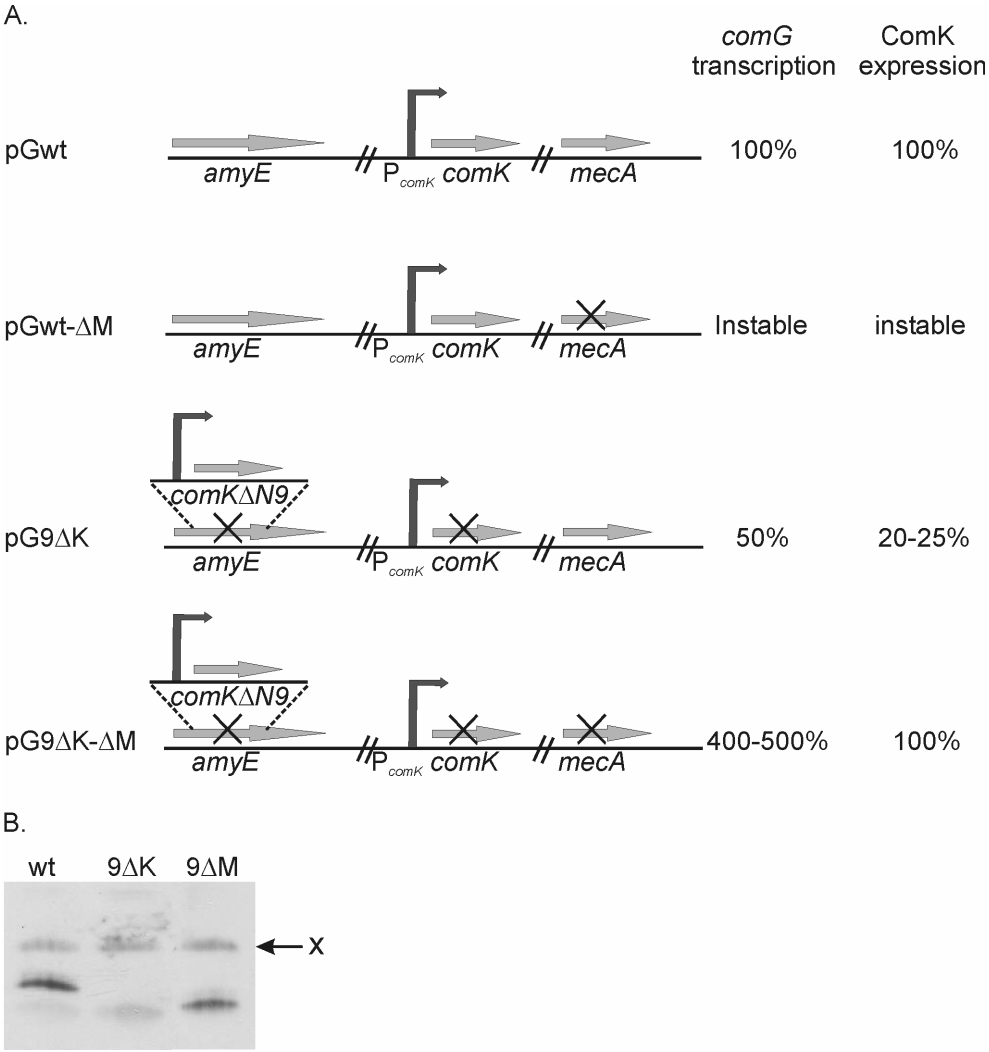


Figure 7. Overview of transcription activation and protein expression for ComKΔN9 and wtComK in different genetic backgrounds
(A). Per strain, the schematic drawing indicates whether the cell contains wtComK or ComKΔN9 (expressed under control of its own promoter in the *amyE* locus) and *mecA*. Deletions are indicated by crosses. *comG*-transcription activation and ComK-expression are shown as percentage of the level observed for *B. subtilis* pGwt. (B). Western blots were performed with a ComK-specific first antibody to demonstrate the level of ComK-expression in the different strains. Wt: pGwt; 9ΔK: pG9ΔK; 9ΔM: pG9ΔK-ΔM. X indicates an aspecific band which serves as a control for the total protein level of the samples.

As is depicted in fig. 7, the levels of ComKΔN9-expression and β-galactosidase activity increased upon introduction of Δ*mecA*, resulting in protein expression comparable to the level observed for wtComK in *B. subtilis* pGwt and a transcription activation of about 4 to 5-fold of the pGwt level. ComKΔN9-expression and transcription activation could not be compared with wtComK under the same conditions, since strain *B. subtilis* pGwt-ΔM turned out to be instable. However, previous research demonstrated a dramatic overexpression of wtComK in a Δ*mecA*-background to levels of several times the wildtype ComK level (Kong and Dubnau, 1994).

Discussion

An intriguing observation concerning the role of ComK in the competence regulatory pathway is that ComK is involved in different types of interaction, i) protein-protein interactions with MecA and with other ComK-proteins in oligomerization and ii) protein-DNA interactions. DNA-binding and MecA-interactions could be exclusive, but at least DNA-binding and oligomerization occur simultaneously. An interesting question is if the small 22.8 kDa ComK-protein contains specific functional domains responsible for one or more of these interactions. *In silico* analyses did not provide insight in the location of functional domains in ComK, although sequence comparisons between ComK-proteins of different bacteria revealed a high degree of conservation throughout the entire protein, with the exception of the N-terminal region (fig. 1). Furthermore, the C-terminal part shows only conservation between subgroups of bacterial species. The present study reports an attempt to elucidate the role of the N- and C-terminal region of ComK in competence regulation by the construction of truncation variants.
For this purpose, a test-system was developed which enabled a maximum dissection of ComK-functions. DNA-binding and oligomerization were investigated *in vitro* with electrophoretic mobility shifts assays (EMSA's), while transcription activation was tested in an *in vivo* system in *L. lactis*, which limited the effects of other regulators, like MecA, on ComK. Using this approach, we demonstrated that *B. subtilis* ComK could be overexpressed with an N- or C-terminal His-tag fusion in *L. lactis*, while retaining transcription activation activity (fig. 2). Furthermore, the DNA-binding and oligomerization behaviour of the His-tagged ComK-proteins was

very similar to that of a control wtComK (fig. 2). However, transcription activation observed from a *comG-lacZ* fusion on pG-wt in *L. lactis* is about 10-fold lower than for pG-wt in *B. subtilis*, although nisin-induced ComK-expression in *L. lactis* should be sufficient to stimulate *comG*-transcription to high levels. It might be that the presence of a His-tag on ComK decreases transcription activation. Indeed, the position of the tag was shown to influence DNA-binding and transcription activation by ComK, as was shown in fig. 2 for His_6-ComK and ComK-His_6. Another possible explanation could be that transcription at the *B. subtilis comG*-promoter is lower because of differences between *L. lactis* RNA-polymerase and *B. subtilis* RNAP. Nevertheless, the fact that *B. subtilis* ComK can activate transcription from the *comG*-promoter in *L. lactis*, provides a suitable test-system for ComK-truncation mutants.
The C-terminal region of ComK is particularly interesting because it displays sequence homology between subgroups of ComK-proteins, suggesting a specific role for this region (fig. 1). C-terminal truncation variants of ComK were overexpressed as N-terminal His-tag fusions in *L. lactis*, demonstrating that transcription activation is completely abolished when 25 amino acids are removed (fig. 3). Interestingly, DNA-binding affinity of His_6-ComKΔC25 is not affected compared to wtComK, while affinity of the 35 amino acids truncation variant His_6-ComKΔC35 is clearly decreased. Although the affinity for the DNA is not affected in His_6-ComKΔC25, the observed shift is smaller (fig. 3) and could be explained by binding of a single dimer instead of a tetramer. Previous research showed that oligomerization of ComK-dimers into a tetramer can only occur when both AT-boxes are located on the same side of the DNA-helix (Hamoen *et al*., 1998). When both boxes are positioned on opposite sides of the helix, ComK is forced to bind as a dimer. By inverting the orientation of the AT-boxes, the possibility that the smaller shift is due to a decrease in protein size of the mutants could be ruled out (fig. 4). Furthermore, the same experiment demonstrated a similar shift for His_6-ComKΔC25 or ΔC35 on both probes, indicating that the smaller shift on the wildtype K-box indeed results from dimer binding. This observation implies that the C-terminal region of ComK is required for tetramerization of the protein, which will affect DNA-bending, another important characteristic of ComK-binding (Hamoen *et al*., 1998). Previous research on the *comG*-promoter showed that the major effect of ComK on transcription activation is on stabilizing RNAP-binding,

likely via facilitating contacts between RNAP and upstream DNA, enabled by DNA-bending (chapter 2). The loss of transcription activation by the C-terminal truncation variants can therefore be explained by a defect in tetramerization and thus DNA-bending.
In addition to a function for the C-terminal region of ComK, this study also investigated the role of the N-terminal part of ComK in the competence regulatory pathway. To our surprise, a truncation of 9 amino acids from the N-terminal end of ComK yielded a hyperactive variant in transcription activation in *L. lactis* (fig. 5). In addition, DNA-binding affinity is higher as well. However, using *L. lactis* as a host, might mask effects of the N-terminal truncation in relation to other regulatory inputs, which are present in *B. subtilis*, but not in *L. lactis*. Compared with the complex competence regulatory pathway in *B. subtilis*, the system in *L. lactis* is incomplete, since it lacks for example MecA. Furthermore, by using an inducible system, potential effects on the autostimulatory loop involved in *comK*-transcription are missed. In order to determine the biological relevance of the N-terminal region of ComK, a ComKΔN9 variant was introduced in the *amyE*-locus of *B. subtilis* under control of the *comK*-promoter. Comparing transcription activation by ComKΔN9 with wtComK demonstrated that, although the absolute level of transcription of *comG-lacZ* is 2-fold lower, ComKΔN9 is relatively more active in transcription activation, since the protein expression level is even 5-6 fold lower than wtComK-expression (fig. 6 and 7). A possible explanation for the decreased expression of ComKΔN9 could be an altered interaction with MecA/ClpCP, resulting in a faster breakdown of ComK. To investigate the interactions with MecA, a *mecA* deletion was introduced, resulting in an increase in protein expression to wildtype level accompanied by a 4-5 fold enhancement of transcription activation by ComKΔN9. However, the proper control experiment could not be performed, since a stable *B. subtilis* Δ*mecA* strain could not be obtained, despite several attempts (fig. 7). Previous research described the construction of such a mutant, although high revertance is reported, which might cause the instability of the strain. However, the reports described a level of wtComK-overexpression that was much higher than the level achieved in this study for ComKΔN9 (Kong and Dubnau, 1994), suggesting that, despite the increase of ComKΔN9-expression in a Δ*mecA*-background, an altered interaction between ComKΔN9 and MecA is not the main reason for the low expression level. As an alter-

native, a *clpC*-deletion was introduced, since ClpC acts on the same proteolytic control system as MecA. Δ*clpC* strains could be obtained in both a *B. subtilis* pGwt and a pG9ΔK background, although again pGwt was more affected by the lack of ClpC than pG9ΔK, as became clear from the slower growth-rate of *B. subtilis* pGwt. In the Δ*clpC* strain pGwt-ΔC, transcription activation by wtComK and its protein expression level increased 3-4 fold compared to *B. subtilis* pGwt. In contrast however, transcription activation by ComKΔN9 increased only slightly to 50-75% of the level of pGwt, which means little or no increase compared to the original ComKΔN9 strain, *B. subtilis* pG9-ΔK (fig. 7). Taken together, these results suggest that the low expression of ComKΔN9 is not due to altered interactions with the MecA/ClpCP proteolytic degradation complex. Other possible explanations for the decreased protein expression of ComKΔN9 could be a reduced transcription activation at the *comK*-promoter or, alternatively, a decreased intrinsic stability of ComKΔN9. The latter possibility is not the most likely, since the protein could be normally overexpressed in *L. lactis*, without displaying protein instability. Since ComKΔN9 was shown to be more active in transcription activation, it might not seem very likely that the reduced protein production is due to a lowered transcription activation of *comK*. However, in this study, transcription activation was tested at the *comG*-promoter, which differs significantly from the *comK*-promoter and might be easier to activate. For example, the *comG*-promoter does not need any other activators than ComK, while at the *comK*-promoter DegU is required to stimulate ComK-activated transcription when ComK-concentrations are low. Furthermore, in the K-box of the *comK*-promoter the spacing between the start positions of both AT-boxes spans four helical turns, compared with three turns in the *comG*-promoter. It might be that this longer spacing requires other abilities of the ComK-protein. If the N-terminal region of ComK would be required for one of these mechanisms, a truncation of this part would result in lower transcription activation at the *comK*-promoter and therefore in a reduced ComK-production. More research is required to determine the exact role of the N-terminal region of ComK in the function of the protein. However, the fact that truncation of the N-terminal region of ComK results in higher transcription activation, but lower protein expression suggests that the presence of this region is a trade-off between optimal transcription activation and a thus far unidentified role in

regulation of transcription, activity or stability of ComK.
This study presents a general overview of the location of two domains affecting ComK-behaviour. Further research is required to determine the exact role of the N-terminal region of ComK in activity and/or stability of the protein. Furthermore, despite the elucidation of the C-terminal region as required for tetramerization, the exact mechanism and the amino acid residues involved remain unknown. However, it is tempting to speculate on possible mechanisms underlying tetramerization. For example, the C-terminal part of ComK contains many charged residues, suggesting that the distribution of negative and positive charges enables electrostatical interactions of this region with the C-terminal part of other ComK-proteins or, alternatively, interactions with other charged regions in the protein. The same mechanism could apply for ComK-proteins in other species, since all ComK-proteins display relatively many charged residues in their C-terminal region.
Functional domains for the other indicated ComK-interactions, like dimerization, DNA-binding and interactions with MecA, have not been identified yet. However, the significantly lower DNA-binding affinity of ComKΔC35, as compared to wtComK and ComKΔC25, may indicate that the C-terminal region of ComK also forms part of the DNA-binding domain. This interpretation is supported by the observation that ComK displays limited sequence homology with the DNA-binding domain of human hSRY, which interestingly is a minor groove binder as well (Werner *et al*., 1995). The region showing the highest homology is slightly touched in ComKΔC35, while most of the homologous region is located further towards the middle segment of ComK and is not affected by any of the truncations. It has been suggested before that the DNA-binding mechanism of ComK could be similar to that of hSRY (Hamoen *et al*., 1998), which would be an interesting starting point for further research. In addition, further research is required to elucidate the domains involved in dimerization and interactions with MecA.

Acknowledgements

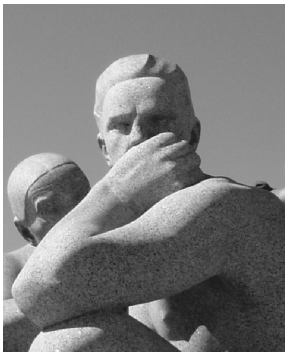
We would like to thank Wiep Klaas Smits and Jan-Willem Veening for helpful discussions. Furthermore, we would like to thank Anton Steen and Rasmus Larsen for help with the first work in *L. lactis*.

Chapter 4

A single, specific thymine mutation in the ComK-binding site severely decreases binding and transcription activation by the competence transcription factor ComK of *Bacillus subtilis*

Kim A. Susanna, Leendert W. Hamoen and Oscar P. Kuipers

This chapter has been submitted for publication



Abstract

The competence transcription factor ComK plays a central role in competence development in *Bacillus subtilis* by activating transcription of the K-regulon. ComK-activated genes are characterized by the presence of a specific sequence to which ComK binds, a K-box, in their upstream DNA region. Each K-box consists of two AT-boxes, with the consensus sequence AAAA-$(N)_5$-TTTT, which are separated by a flexible spacing resulting in two, three or four helical turns between the starting nucleotides of the repeating AT-box units. In this study, the effects of potential determinants of ComK-regulation in K-boxes were investigated by testing transcription activation and DNA-binding affinity on altered K-boxes, which were either mutated in the spacing between the AT-boxes or in the consensus sequence of the AT-boxes.
The most striking result demonstrates the importance of the second thymine base in the AT-boxes. Mutation of this T into a guanine resulted in a 3-fold reduction in transcription activation and DNA-binding by ComK. Transcription activation, as well as DNA-binding, was almost completely abolished when both AT-boxes contained a T_2 to G mutation. This result was corroborated by *in silico* analyses, demonstrating that a combination of mutations at the T_2 positions of both AT-boxes is not found among ComK-activated K-boxes, indicating that at least one consensus T_2-position is required to maintain a functional K-box. The results suggest an important structural role of T_2 in ComK-binding, probably by its specific position in the minor groove of the DNA.

Introduction

The development of genetic or natural competence is an adaptation process that enables bacterial cells to take up and integrate exogenous DNA into their genome. This phenomenon has been studied extensively in the Gram-positive soil bacterium *Bacillus subtilis*. Competence development depends on the presence of a key regulatory protein, *i. e.* the competence transcription factor ComK. During exponential growth, the presence of free ComK in the cell is prevented by both transcriptional and post-translational control. Transcription of *comK* is repressed by binding of AbrB, CodY and Rok to the *comK*-promoter (Hamoen *et al*., 2003a; Hoa *et al*., 2002; Serror and Sonenshein, 1996), while any ComK that is produced is trapped by MecA and targeted for proteolytic degradation by ClpCP (Turgay *et al*., 1998). At the end of the exponential growth, increased cell densities are sensed and interpreted by a quorum sensing pathway. This results in the production of ComS, a small protein that can liberate ComK from the proteolytic complex (D'Souza *et al*., 1994; Hamoen *et al*., 1995; Solomon *et al*., 1995). At the same time, the repressing effect of CodY and AbrB is relieved in response to nutrient limitation. Free ComK can bind to its own promoter and activate gene transcriptt-ion (Van Sinderen and Venema, 1994b; Van Sinderen *et al*., 1995). Furthermore, ComK can repress *rok*-transcription by binding to the *rok*-promoter, thereby relieving the repression by Rok of *comK*-transcription (Hoa *et al*., 2002). The ensuing autostimulation of *comK*-transcription results in a rapid increase of the ComK-concentration in the cell. In addition to its own gene, ComK stimulates transcription of the K-regulon, *e. g.* the late competence genes, encoding the DNA-binding, -uptake and -integration machinery (review: Dubnau, 1993; Hamoen *et al*., 2003b).
ComK activates transcription by binding to specific sequences, so-called K-boxes, in the upstream region of ComK-activated genes (fig. 1). Each K-box consists of two AT-boxes with the consensus sequence AAAA-$(N)_5$-TTTT. The boxes are separated by a spacing of a discrete number of helical turns, which positions both boxes on the same side of the DNA-helix. Based on the spacing, which can have a length of two, three or four helical turns when calculated between the first A's of both AT-boxes, K-boxes are divided in class I, class II or class III promoters, respectively. Functional ComK is thought to act as a tetramer composed of two dimers,

which each bind to an AT-box. Tetramerization of the two dimers is accompanied by DNA-bending (Hamoen *et al*., 1998). Chapter 2 has demonstrated that the major role of ComK in transcription activation at the *comG*-promoter is on stabilizing RNA-polymerase binding, probably by facilitating contacts between RNAP and the upstream DNA through bending of the promoter region.

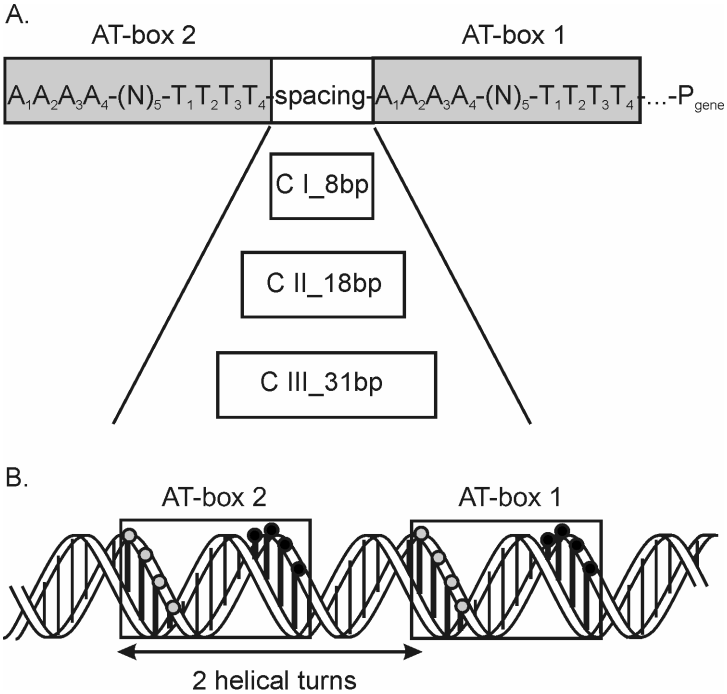


Figure 1. Overview of the ComK-binding site, K-box, in *B. subtilis*
(A). Schematic representation of a K-box. A K-box consists of two AT-boxes (grey) with the consensus sequence AAAA-$(N)_5$-TTTT. In this study, the AT-box closest to the -35 is indicated as AT-box 1, while the box further upstream is called AT-box 2. In each AT-box, the positions of the 4 A's and T's are indicated with numbers in subscript. The three classes of K-boxes are indicated with their respective spacings. The average AT-content of the spacing region is over 60%. (B). Helical representation of a K-box. A class I K-box with two turns spacing is shown in a helical projection. The dark vertical bars represent the AT-basepairs in the AT-boxes. Grey circles show the $(A)_4$- and black circles the $(T)_4$-stretches.

Recently, three transcriptome studies were reported to define the ComK-regulon in *B. subtilis* (Berka *et al*., 2002; Hamoen *et al*., 2002; Ogura *et al*., 2002). *In silico* analyses demonstrated the presence of over a 1000 putative K-boxes in the genome (one-

third in intergenic regions) with up to three basepair deviations from the consensus K-box sequence with a maximum of two per AT-box. Previous reports demonstrated that these characteristics can still yield functional K-boxes (Hamoen *et al*., 1998). Based on the combined results of the three transcriptome studies, it was shown that, under laboratory conditions, only 8% of these genes were indeed regulated by ComK, indicating that the sole presence of a K-box is not enough to predict regulation by ComK (Hamoen *et al*., 2002). Probably, additional elements in or near K-boxes are important for activity of the boxes in regulation by ComK.
As is depicted in fig. 1, different elements can be distinguished in a K-box, like the spacing between the AT-boxes and the consensus sequence of these boxes. In principle, these characteristics could serve as additional critical elements to determine whether a K-box is regulated and, when regulated, to which level ComK stimulates transcription at this particular box. In the present study, a search for critical determinants in a K-box was performed by investigating the effects of alterations in the length and GC-content of the spacing between the AT-boxes. Furthermore, the effects of point-mutations in the consensus sequence of the AT-boxes were determined using the K-box of the *comG*-operon as a model, as well as an idealized K-box. This K-box was chosen because transcription of the *comG*-operon is among the highest ComK-activated transcriptions and requires only ComK, minimizing the chance of interference of other regulatory proteins (Hamoen *et al*., 1998; chapter 2). The effects on regulation by ComK were investigated by monitoring transcription activation *in vivo,* using β-galactosidase assays and by determining DNA-binding affinity of ComK for wildtype and mutant K-boxes *in vitro* by electrophoretic mobility shift assays. The results show the importance of the second thymine basepair in an AT-box for regulation of the K-box by ComK, suggesting that this position is crucial for ComK-DNA interactions.

Materials and Methods

Bacterial strains, media and growth conditions
The bacterial strains used in this study are listed in table 1. *B. subtilis* strains were grown in minimal medium (Venema *et al*., 1965) supplemented with erythromycine (5 μg/ml). *E. coli* strains were grown in TY-medium (Sambrook *et al*., 1989) + 0.2% glucose and ampicillin (100 μg

/ml). MBP-ComK overexpression was induced with 0.3 mM IPTG (pMal protein fusion and purification system, NEB). *L. lactis* strains were grown in 2-fold diluted M17 based medium (Difco), supplemented with 0.5% glucose (GM17) and, if required, erythromycine (4 μg/ml).

Table 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant feature	Reference
Strains		
B. subtilis 168	*trpC2*	Anagnostopoulos and Spizizen, 1961
L. lactis NZ9000	MG1363 *pepN::nisRK*	Kuipers *et al.,* 1998
E. coli XL1blue	*endA1 gyrA96 thi hsdR17 (rK-mK-) supE44 relA1 lac/F' proAB laclq lacZΔM15 Tn10*	Stratagene
Plasmids		
pNZ8048	Cm^r, inducible expression vector containing the *nisA*-promoter	Kuipers *et al*., 1998
pMal-ComK	Amp^r, containing *malE-comK* fusion	Chapter 3
pILORI4	Ery^r, pIL252 carrying the multiple cloning site and promoterless *lacZ* of pORI13	Larsen *et al.,* 2004
pILORI-G	Ery^r, *comG*-promoter-*lacZ* fusion	This work
pILORI-C1	Ery^r, contains idealized class I K-box	This work
pILORI-C2	Ery^r, contains idealized class II K-box	This work
pILORI-C3	Ery^r, contains idealized class III K-box	This work
pG-b1/2P-SpGC-middle	Ery^r, perfect *comG*-promoter K-box, with 60% overall GC-content in spacing, GC's in middle of spacing	This work
pG-b1/2P-SpGC-end	Ery^r, perfect *comG*-promoter K-box, with 60% overall GC-content in spacing, GC's at ends of spacing	This work
pG-wt	Ery^r, wildtype common K-box of *comG*-promoter	This work
pG-perfect	Ery^r, wildtype K-box of *comG*-promoter mutated to perfect K-box sequence	This work
pG-b2P-b1A_x-G	Ery^r, K-box with perfect AT-box 1 + A_xG mutation in AT-box 2 (x = 1, 2, 3 or 4)	This work
pG-b2P-b1T_x-G	Ery^r, K-box with perfect AT-box 1 + T_xG mutation in AT-box 2 (x = 1, 2, 3 or 4)	This work
pG-b2A_2G-b1A_xG	Ery^r, K-box with A_2-G in AT-box 1 + A_xG mutation in AT-box 2 (x= 1 or 3)	This work
pG-b2A_2G-b1T_xG	Ery^r, K-box with A_2-G in AT-box 1 + T_xG mutation in AT-box 2 (x = 2, 3 or 4)	This work
pG-b2T_2G-b1P	Ery^r, T_2G mutation in AT-box 2 as single bp change in a perfect K-box	This work
pG-b2T_2G-b1T_2G	Ery^r, T_2-G mutations in both AT-boxes	This work

DNA techniques, materials and transformations

Standard molecular biology methods were used as described (Ausubel *et al*., 1998). Enzymes were purchased from Roche, New England Biolabs or Pharmacia. Radiolabeled nucleotides were obtained from Amersham.

For plasmid isolations and PCR product purifications, the High Pure Plasmid Isolation kit and the High Pure PCR Product Purification kit, respectively, were used (Roche). *B. subtilis* was transformed as described by Anagnostopoulos and Spizizen (1961). Transformation of *L. lactis* NZ9000 was achieved by electroporation using a Gene pulser (BioRad Laboratories, Richmond, Calif.) as described by Leenhouts and Venema (1993).

PCR amplifications and plasmid constructions

PCR reactions were performed as described (Innes and Gelfand, 1990). For amplification of PCR-products, Pwo or Expand DNA-polymerase were used (both from Roche). Unless specified otherwise, chromosomal DNA of *B. subtilis* 168 was used as a template. Plasmids and primers used in this study are listed in table 1 and table 2, respectively.

To investigate the effect of spacing length on regulation by ComK, synthetic DNA sequences were designed, containing an idealized K-box. The K-box contained the perfect AT-boxes found in the promoters of *addAB* (AT-box 1) and *comG* (AT-box 2). The spacing sequence consisted of random DNA with a similar AT-content as *B. subtilis* intergenic regions: 8 (CCCAAGGC), 18 (CTCTACCCAAGGCAGT GC) or 31 (CTCTACCAGCATA CCAAGGTATGGTCAGTGC) bp of this random sequence were inserted between the AT-boxes, corresponding with the lengths found in natural K-boxes. The DNA upstream of the boxes contained 20 bp of DNA located on this position relative to the AT-boxes in the *comC*-promoter, while the DNA directly downstream of the boxes arises from the equivalent position in the *comG*-promoter. These oligo DNA fragments were used as a template for PCR amplifications using primers oligo-start-*Eco*RI and oligo-end-*Bam*HI and, after digestion with *Eco*RI and *Bam*HI, cloned into *Eco*RI/*Bam*HI digested plasmid pILORI-G, resulting in plasmids pILORI-C1, pILORI-C2 and pILORI-C3, as representatives of a class I, II and III K-box, respectively. pILORI-G contains a *comG*-promoter-*lacZ* fusion, constructed as follows: the promoter region was amplified with primers comG-start-*Bam*HI and comG-end-*Xba*I. After digestion with *Bam*HI and *Xba*I, the PCR-fragment was cloned into *Bam*HI/*Xba*I digested plasmid pILORI4, upstream of the promoter-less *lacZ* gene.

Plasmids pG-b1+2P-SpGC-middle and pG-b1+2P-SpGC-end were constructed by amplification of the *comG*-promoter region and K-box with primers comG-b2P-SpGC-middle or comG-b2P-SpGC-end, respectively, combined with primer comG-end-*Xba*I. The PCR-products were digested with *Eco*RI and *Xba*I and ligated into *Eco*RI/*Xba*I digested pILORI4. In both plasmids, the GC-content of the spacing is increased to 60%.

Table 2. Primers used for the construction of different K-box mutants
For pILORI-G and pILORI-Cx, the indicated primers were used. For mutations in AT-boxes or spacing sequence, the indicated forward primer was used with primer comG-end-*Xba*I. Annealing sequences are underlined and mutation(s) indicated in bold. Restriction sites are shown in italic.

Plasmids	**Primer**	**Sequence**
Basic		
pILORI-G	comG-start-BamHI	GATC*GGATCC*TTGATTACCTTTTCTTCTTTTTCTACAA TATGCG
	comG-end-XbaI	GATC*TCTAGA*TTATGCCTCTTCAATCAAGTTTTTGC
pILORI-Cx	Oligo-start-EcoRI	GATC*GAATTC*CATCCGGCTCCGGCAGAATC
	Oligo-end-BamHI	GATC*GGATCC*AAAACGGCCTTTTGC
Mutations	**Forward primer**	**Sequence**
pG-wt	comG-AT2-EcoRI	GATC*GAATTC*AGAATTGGTTTTTCAGCATATAACATC TCAC
pG-perfect	comG-perfect	GATC*GAATTC*A**A**AATTGGTTTTTCAGCATATAACATC TC
pG-b2P-b1A1G	comG-b1A1G	GATC*GAATTC*AAAATTGGTTTTTCAGCATATAACATCT CAC**G**AAATC
pG-b2P-b1A2G	comG-b1A2G	GATC*GAATTC*AAAATTGGTTTTTCAGCATATAACATCT CACA**G**AATC
pG-b2P-b1A3G	comG-b1A3G	GATC*GAATTC*AAAATTGGTTTTTCAGCATATAACATCT CACAA**G**ATCACG
pG-b2P-b1A4G	comG-b1A3G	GATC*GAATTC*AAAATTGGTTTTTCAGCATATAACATCT CACAAA**G**TCACG
pG-b2P-b1T1G	comG-b1T1G	GATC*GAATTC*AAAATTGGTTTTTCAGCATATAACATCT CACAAAATCACG**G**TTTCCCTG
pG-b2P-b1T2G	comG-b1T2G	GATC*GAATTC*AAAATTGGTTTTTCAGCATATAACATCT CACAAAATCACGT**G**TTCCCTG
pG-b2P-b1T3G	comG-b1T3G	GATC*GAATTC*AAAATTGGTTTTTCAGCATATAACATCT CACAAAATCACGTT**G**TCCCTG
pG-b2P-b1T4G	comG-b1T4G	GATC*GAATTC*AAAATTGGTTTTTCAGCATATAACATCT CACAAAATCACGTTT**G**CCCTG
pG-b2T2G-b1P	comG-b2T2G	GATC*GAATTC*AAAATTGGTT**G**TTCAGCATATAACATC TC
pG-b1+2P-spGC-middle	comG-b2P-spGC-middle	GATC*GAATTC*AAAATTGGTTTTTCAGCA**CC**T**GG**CATC TCACAAAATCACG
pG-b1+2P-spGC-end	comG-b2P-spGC-end	GATC*GAATTC*AAAATTGGTTTTTC**C**GC**G**TATAAC**G**TC TC**G**CAAAATCACG

To investigate the effect of point-mutations in the AT-box sequences, *comG-lacZ* fusions were constructed. The wildtype *comG*-promoter fragment containing the common K-box was amplified with primers comG-AT2-*Eco*RI and comG-end-*Xba*I. To obtain an idealized K-box in the *comG*-promoter, the guanine at position 2 in the A-stretch of AT-box 2 was mutated into an adenine by amplifying the *comG*-promoter fragment with primers comG-perfect and comG-end-*Xba*I. In both cases, the PCR-products were digested with *Eco*RI and *Xba*I and ligated into *Eco*RI

/*Xba*I digested pILORI4, resulting in plasmids pG-wt and pG-perfect, respectively. Using the same strategy, point-mutations were introduced in the AT-boxes. Plasmids were constructed using the primer combinations listed in table 2. Single mutations were made in AT-box 1, while AT-box 2 remained perfect. The resulting plasmids were named pG-A_xG or pG-T_xG, with x referring to the position in the A or T stretch of AT-box 1 which is mutated into a G. Some single mutations were combined with the A_2-G mutation in AT-box 2 that is present in the wildtype K-box of the *comG*-promoter. In these cases, plasmids containing the single mutations were used as template for PCR amplification with primers comG-AT2 and comG-end-*Xba*I, yielding a *comG*-promoter fragment containing two mutations in the K-box. The created plasmids are named pG-b2A_2G-b1A_xG or pG-b2A_2G-b1T_xG, again with the x referring to the mutated position. Plasmid pG-b2T_2G-b1P with a T_2-G mutation in AT-box 2 was constructed by amplification of the K-box region with primers comG-b2T_2G and comG-end-*Xba*I. The same primers were used to construct plasmid pG-b1T_2G-b2T_2G, combining the T_2-G mutation in AT-box 2 with a T_2-G mutation in AT-box 1, using plasmid pG-b1T_2G as template. For all plasmids, cloning was performed in *L. lactis* NZ9000. Plasmids were checked by sequencing and transformed into *B. subtilis* 168.

Transcription activation assays

Transcription activation by ComK on wildtype and mutant K-boxes upstream of the *comG*-promoter was tested in *B. subtilis* using fusions with *lacZ* as a reporter. Cultures were grown in minimal medium to stimulate competence development and samples were taken from transition point until 3 to 4 hours into the stationary growth-phase, with one-hour intervals. Samples were analysed for β-galactosidase activity as described by Israelsen *et al.* (1995) and for protein expression levels with SDS-PAGE (Laemmli *et al*., 1970) and Western blots (Towbin *et al*., 1979). ComK was detected with a ComK-specific first antibody (Van Sinderen and Venema, 1994b) and an anti-rabbit horseradish peroxidase conjugated secondary antibody (Amersham) and visualized by chemiluminescent detection using the ECL Western blotting analysis system (Amersham).

Electrophoretic mobility shift assays

Electrophoretic mobility shift assays (EMSA's) were performed essentially as described (Van Sinderen *et al*., 1995). For this purpose, ComK was purified according to the method of Hamoen *et al*. (1998). dsDNA probes were amplified with PCR, using the same primer combinations as used for the construction of the plasmids (table 2). Probes for deter-

mining the effect of spacing length were amplified with primers oligo-start-*Eco*RI and comG-end-*Xba*I. Probes were end-labeled with T4 polynucleotide kinase using [γ-^{32}P]-ATP. Proteins and DNA probes were premixed on ice in 20 µl binding buffer (20 mM Tris-HCl (pH 8.0), 5 mM $MgCl_2$, 100 mM KCl, 0.5 mM dithiotreitol, 0.05 mg/ml poly[d(I-C)], 0.05 mg/ml BSA and 8.7% glycerol). Samples were incubated for 15 minutes at 37ºC to allow protein-DNA complexes to form. 15 µl of each sample was loaded on a nondenaturing 6% polyacrylamide gel. Gels were run in TBE buffer (89 mM Tris, 89 mM boric acid and 2 mM EDTA) at 100 V, dried, and autoradiographed using phosphor-screens. Protein shifts were quantified using the Quantity one software package.

Results

Class III K-boxes are less active than class I and II K-boxes
In the *B. subtilis* genome, three classes of K-boxes can be distinguished based on the length of the spacing between the AT-boxes, which amounts to either 8, 18 or 31 basepairs, corresponding with two, three or four helical turns separating the starting positions of the AT-boxes for class I, II or III promoters, respectively. These spacings position both AT-boxes on the same side of the DNA-helix, thereby enabling interactions between the ComK-dimers that each bind to an AT-box (Hamoen *et al*., 1998). To investigate the effect of the spacing length on ComK-binding and transcription activation, synthetic DNA fragments were developed and cloned upstream of a *comG*-promoter-*lacZ* fusion, thereby replacing the original K-box of *comG*. These DNA fragments contained 'perfect' AT-boxes, separated by 8, 18 or 31 random basepairs, resulting in representatives of a class I, II or III promoter, respectively. β-galactosidase assays showed that ComK activates transcription of all three classes, but that the level of transcription is lower with a class III K-box (fig. 2A), while there is little or no difference in transcription activation between a class I and a class II promoter. The same is seen in DNA-binding by ComK. Using electrophoretic mobility shifts assays (EMSA's), ComK-binding to the DNA of a K-box was investigated for a range of ComK-concentrations (fig. 2B). The percentage of shifted probe was determined for each ComK-DNA combination, demonstrating that the affinity of ComK for a K-box is the lowest when the spacing spans four helical turns, while there is only a slight difference between a

class I and a class II K-box (fig. 2C). Binding of ComK at a class III K-box can ultimately result in shift levels comparable to shifts at a class I or II K-box, but only at higher ComK-concentrations.
As control experiments, Western blots and plasmid isolations were performed to check the level of ComK and the plasmid content for the different strains in all experiments. These assays demonstrated that ComK-production and plasmid content are comparable for all strains tested (results not shown), indicating that differences in transcription activation are due to the mutations in the K-boxes and not to variation in ComK expression and/or plasmid copy numbers between the strains.

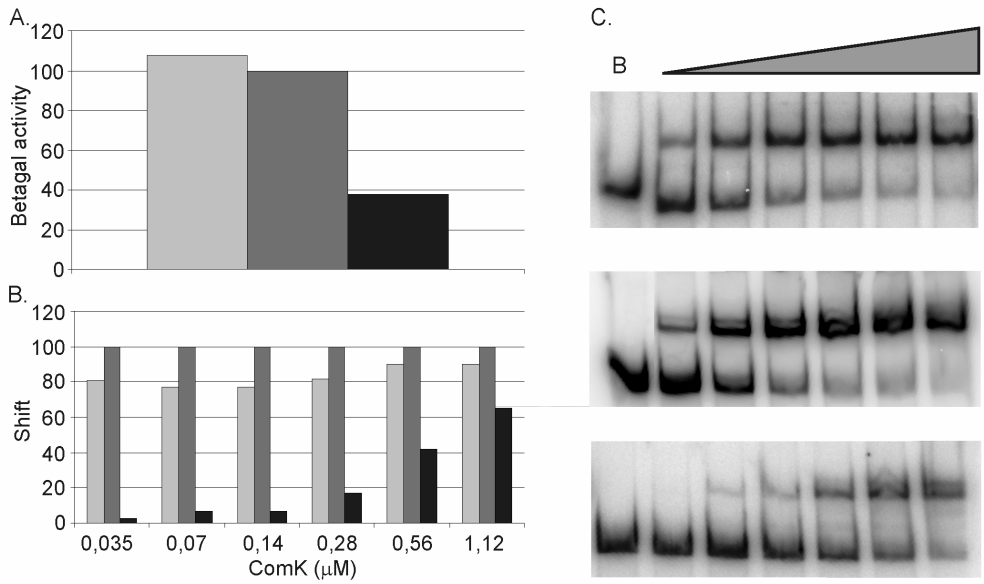


Figure 2. Effect of the length of the spacing on regulation of a K-box by ComK (A). β-galactosidase assays were performed to investigate transcription activation by ComK on perfect K-boxes with different spacing lengths, representing a class I, II or III promoter. The sample for the depicted β-galactosidase assay was taken after 4 hours in the stationary growth phase. β-galactosidase activities are represented as percentages relative to the activity for the class II K-box. (B) and (C). EMSA's were performed to determine DNA-binding by ComK to the three classes of K-boxes. Binding was tested for different ComK-concentrations. (B). ComK-shifts were quantified and represented as the percentage of the shift seen for the class II K-box for every ComK-concentration. Light grey: class I K-box; middle grey: class II K-box; black: class III K-box. (C). Concentration bar: ComK-concentrations increasing with two-fold increments from 0.035 to 1.12 μM. Upper picture: class I K-box; middle: class II K-box; lower: class III K-box.

The spacing GC-content does not affect regulation by ComK

B. subtilis is an AT-rich bacterium, with an average content of about 60% AT-basepairs in the genome sequence. Also the DNA of the spacing region in a K-box is relatively AT-rich. Binding of ComK to a K-box induces a bend in the DNA, which is required for transcription activation, as was shown at the promoter of *comG* (chapter 2). In general, AT-basepairs are regarded as more flexible in bending than the somewhat more rigid GC's (Koo *et al*., 1986; Perez-Martin *et al*., 1994), suggesting that the AT-content of the spacing region might be an important determinant for regulation by ComK. To investigate this possibility, the AT-content was reduced from 60% to 40%, using the *comG*-promoter K-box as a model. In order to focus only on the influence of the spacing sequence, the AT-boxes were changed into a perfect consensus. Furthermore, it can be imagined that the position of the GC-basepairs along the spacing affects bending abilities and thereby regulation by ComK. To take this into account, two constructs were tested: one in which the ends of the spacing were enriched in GC-content and the other in which the AT-GC substitutions were positioned in the middle of the spacing. Surprisingly, these changes in the spacing sequence did not significantly affect transcription activation by ComK. β-galactosidase studies showed only a slight increase (110% of control level) of transcription when the spacing middle was enriched in GC-basepairs and a very slight decrease (95% of the control) when the GC's were introduced at the ends of the spacing. In addition, EMSA's demonstrated comparable binding of ComK to control K-boxes and spacing mutant K-boxes (results not shown).

Position T_2 in AT-box 1 is critical for activation by ComK

As described above, the characteristics of the spacing region do not determine whether a K-box is active in regulation, but rather to which extent the box is activated. In search for possible other features of a K-box that do determine whether and to which level transcription is regulated from a K-box, the influence of mutations in the consensus sequence of the AT-boxes was investigated. To determine the effect of a single basepair mutation on ComK-regulation, a clean test system was developed using the class II promoter of the *comG*-operon as a model. In previous research, this promoter was used to investigate the role of ComK in transcriptt ion activation, since it is among the highest ComK-activated pro-

moters, and transcription activation is only dependent on ComK (chapter 2). The wildtype *comG*-promoter contains an additional AT-box at two helical turns upstream of the start of the common K-box. Since the goal of this research is to focus on determinants of general importance for ComK-regulation, the extra upstream AT-box was omitted and a *comG*-promoter containing only the common K-box was fused with the *lacZ*-gene in plasmid pG-wt. The K-box of *comG* has one basepair deviation from the consensus sequence, *i. e.* an A to G mutation at position 2 in AT-box 2. This mutation was repaired in plasmid pG-perfect, creating a perfect, idealized K-box. As shown in fig. 3A, ComK activated transcription at both promoters to comparable levels. Furthermore, ComK showed similar DNA-binding on both K-boxes (fig. 3B).

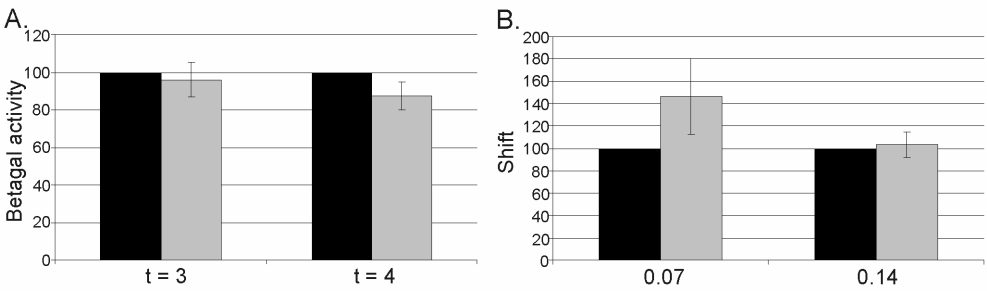


Figure 3. Comparison of wildtype *comG*-promoter K-box and a perfect K-box
(A) β-galactosidase assays were used to determine the levels of transcription activation by ComK at the wildtype *comG*-promoter K-box and a perfect K-box. Samples were taken from transition point until 4 hours into stationary growth. β-galactosidase activities at the last two timepoints are depicted, indicating little difference between both K-boxes. (B) EMSA's were used to test DNA-binding by ComK at both K-boxes. Binding was determined for different concentrations. Shifts were quantified and represented as percentage of the shift at the wildtype K-box for 0.07 and 0.14 μM ComK. Black: wildtype K-box; grey: perfect K-box.

To investigate the effect of single basepair mutations in an AT-box on regulation by ComK, the idealized K-box of the *comG*-promoter was subjected to a systematic, box-scanning mutagenesis, introducing single G's at each position in AT-box 1. Mutants were tested for transcription activation by β-galactosidase assays. As shown in fig. 4A, three groups can be distinguished: i) a group with comparable or slightly higher transcription than pG-perfect (A_2, A_3, A_4, T_3), ii) an intermediate group (A_1, T_1 and T_4) and iii) a low level (T_2), which is 3-fold reduced compared to pG-perfect.

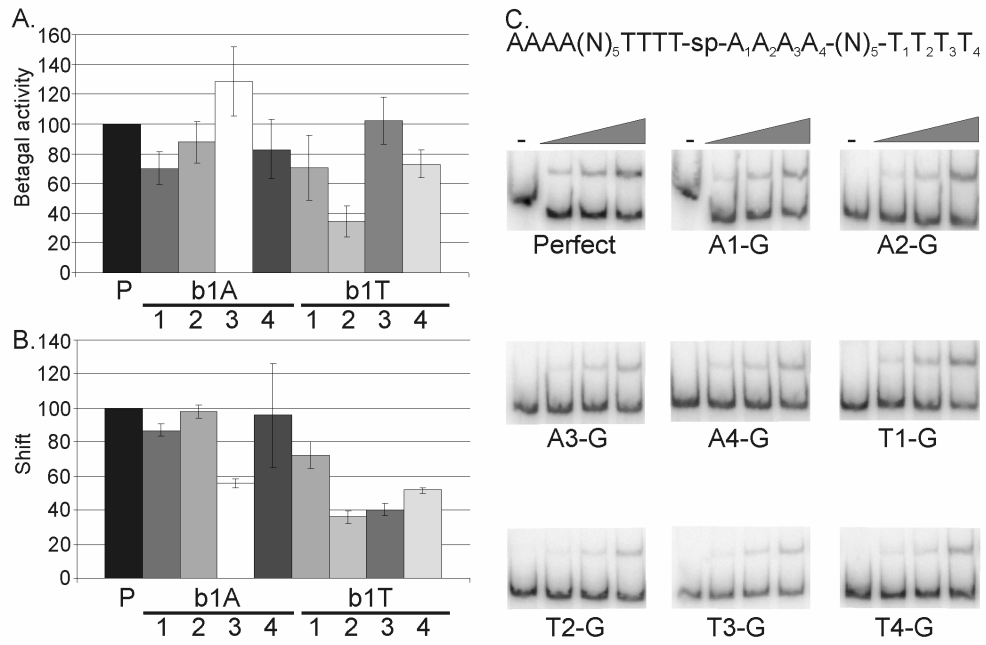


Figure 4. Effect of mutations on transcription activation and binding by ComK
(A) β-galactosidase assays were used to determine the effects of single basepair mutations on transcription activation by ComK *in vivo*. β-galactosidase activities are represented as percentage of activity relative to the level for the perfect K-box in the same experiment. Activities after 4 hours of stationary growth are depicted. Error bars representing standard deviations over the average of three experiments are included for all mutations. Legend A and B: P: perfect K-box; b1A or b1T: numbers 1 – 4 refer to the sequence above (of the perfect K-box) and indicate the position substituted by a G in the A- or T-stretch of AT-box 1. (B) and (C) To determine the effect of the mutations on ComK-binding, EMSA's were performed. Shifts for mutants were quantified and represented as percentages relative to the shifts for the perfect K-box. Standard deviations determined over the average of two experiments are depicted for the mutants. Quantifications are shown for the reaction with 0.07 μM ComK. In each panel in C, a blanc (-) and ComK-concentrations increasing with two-fold increments from 0.018 to 0.07 μM are shown (conc. bar). Perfect: perfect K-box; A_xG or T_xG: A-G or T-G mutation on position x in the A- or T-stretch of AT-box 1.

In addition to affecting transcription activation, single mutations in a K-box also affect DNA-binding by ComK (fig. 4B and 4C). The binding-effects for each position do not always fully correlate with the effects seen on gene transcription, *e. g.* position A_3, for which

transcription and ComK-binding are somewhat differently affected. Except for the T_1-position, all T-stretch mutations show a clear reduction in DNA-binding by ComK. Like the effect of the mutation on transcription activation, a T to G substitution on position T_2 shows the largest effect on ComK-binding, indicating that this basepair position is most important for ComK-DNA interactions.

Transcription and binding on double mutants in a K-box

For further investigations of the importance of the consensus sequence of the AT-boxes, double mutants of a K-box were created. For this purpose, single mutations in the investigated AT-box 1, representing a wildtype, an intermediate and a low level of transcription activation, were combined with the natural mutation that is found in the wildtype *comG*-promoter K-box, *i. e.* A_2-G in AT-box 2. The double mutants were tested for transcription activation and ComK-binding and compared with the wildtype *comG*-promoter, pG-wt. Again, three levels of transcription activation could be distinguished: 60-70% of the wildtype level for A_1 and T_3, intermediate transcription for A_3 and a low level for mutations at positions T_2 and T_4 in AT-box 1, combined with the A_2-G mutation in AT-box 2 (fig. 5A).

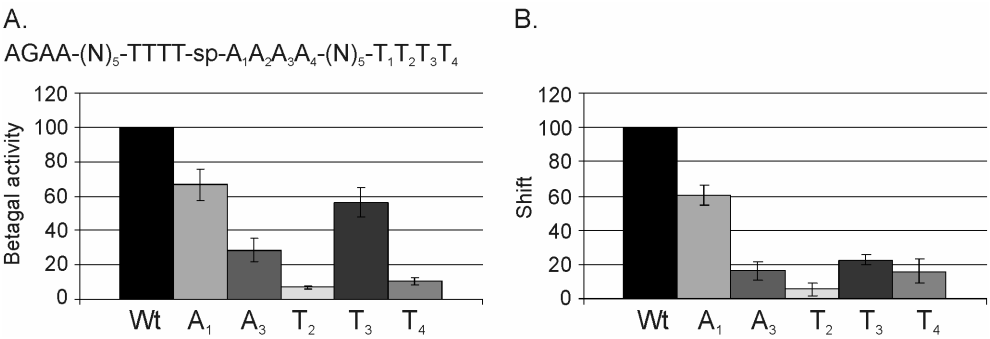


Figure 5. Effect of double mutations on transcription activation and binding
(A) Single mutants in AT-box 1 were combined with the natural A_2-G mutation in AT-box 2 as in the wildtype *comG* K-box (sequence above). β-galactosidase activities and standard deviations are determined as in fig. 4, relative to the *comG*-promoter K-box. (B) EMSA's and quantifications were performed as in fig. 4. The wildtype *comG*-promoter K-box was used as control. Results are shown for the binding reaction with 0.07 μM ComK. wt: wildtype *comG*-promoter K-box; A_x or T_x: x is the position that is substituted by a G in the A- or T-stretch, respectively, of AT-box 1, in addition to the A_2-G substitution in AT-box 2.

Similar effects could be seen on ComK-binding, although these effects seem to be somewhat stronger than on transcription (fig. 5B). This is especially clear for the combination of A_2-G in AT-box 2 with T_3-G in AT-box 1, where DNA-binding affinity is only near 20% of the wildtype level whereas transcription activation is still almost 60% compared with wildtype. Apparently, a decrease in binding affinity does not necessarily result in a similar reduction of transcription activation.

Mutation of both T_2's almost completely inactivates a K-box
As mentioned in the introduction, the *B. subtilis* genome contains over a 1000 K-boxes, with a wide variety of deviations from the consensus sequence. Because of this natural variation, big effects of single mutations in a K-box were not expected. However, as shown in fig. 4, a mutation at position T_2 in AT-box 1, severely disturbs both transcription activation and DNA-binding by ComK. For further investigation of the importance of this position for ComK-DNA interactions, K-boxes with mutated T_2-positions in AT-box 1, AT-box 2 or both AT-boxes were tested for DNA-binding and transcription activation by ComK. As is shown in fig. 6, a T_2-G mutation in AT-box 2 clearly reduces transcription activation and DNA-binding by ComK, compared to wildtype or pG-perfect levels, but reduction is slightly less than with the same mutation in AT-box 1. A combination of mutations at both T_2-positions however, leads to an almost complete loss of transcription activation and DNA-binding by ComK.

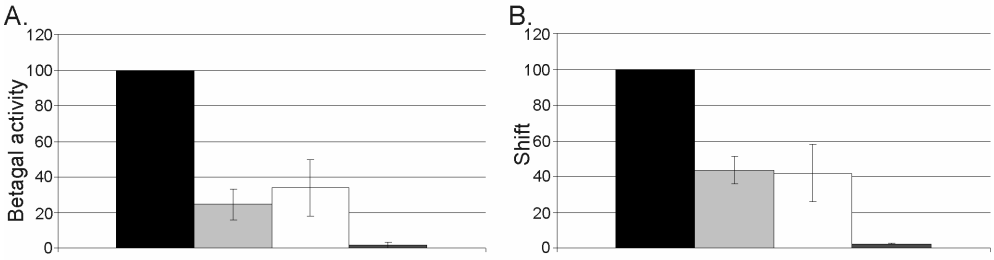


Figure 6. Importance of T_2-positions in both AT-boxes for regulation by ComK (A). β-galactosidase assays were performed as before. Activities are represented relative to the perfect K-box. Error bars are determined as for fig. 4 and 5. (B). EMSA's were performed as before. Shifts are represented relative to the perfect K-box. Black: perfect K-box; light grey: T_2-G in AT-box 1, perfect AT-box 2; white: T_2-G in AT-box 2, perfect AT-box 1; dark grey: T_2-G in both AT-boxes.

In silico analyses of ComK-regulated K-boxes

In 2002, three transcriptome studies on the ComK-regulon were reported, providing an overview of natural transcription activation levels at K-boxes with various deviations from the consensus sequence (Berka *et al*., 2002; Hamoen *et al*., 2002; Ogura *et al*., 2002). In this study, those ComK-regulated genes that were found in all three studies were grouped, ranging from strong to weaker activation by ComK. Only unique K-boxes were included, by taking into account only the highest fold activated gene from an operon. The resulting list contained 88 ComK-regulated genes and was used as a reference provided by *B. subtilis* itself to determine the occurrence of mutations at each AT-box position. A weight matrix based on these ComK-regulated K-boxes shows a wide variety of mutations throughout the complete K-box (fig. 7A), correlating with the previous knowledge that K-boxes with up to three deviations from the consensus sequence can still be good targets for regulation by ComK. 11 out of these 88 ComK-activated genes were shown to contain a T-N substitution on position 2 in AT-box 1. A weight matrix based on these K-boxes clearly demonstrated that a concomitant mutation on the other T_2-position in AT-box 2 is never seen among ComK-regulated boxes (fig. 7B).

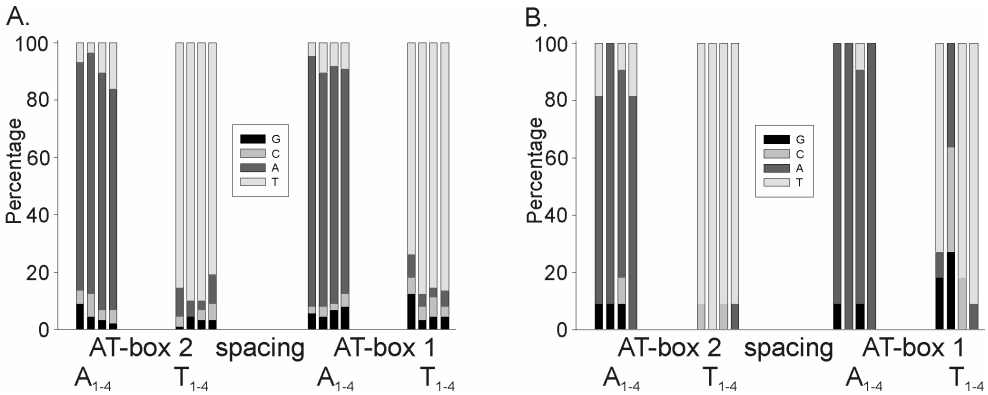


Figure 7. Weight matrices showing natural variation ComK-regulated K-boxes (A). Weight matrix based on the sequences of 88 ComK-regulated K-boxes. In order to align the AT-boxes for the three classes of K-boxes, the spacing was in all cases set to the same length. (B) Weight matrix based on the sequence of 11 out of the 88 ComK-regulated K-boxes, which contain a T-N substitution on position 2 in AT-box 1. In these mutant K-boxes, the T_2-position in AT-box 2 is strictly conserved.

When on the other hand, non-ComK-regulated boxes with a mutation on the T_2-position of either AT-box 1 or 2 were used to construct a weight matrix, mutations on the T_2-position in the other box could be observed (results not shown). These results show a very good correlation with the experimental data from this study, strongly suggesting that mutations occuring simultaneously at both T_2-positions in one K-box result in inactivation of this box.

Discussion

The competence transcription factor ComK regulates gene transcription by binding to specific sequences, the so-called K-boxes, located upstream of promoters of regulated genes. Previous research determined a defined consensus sequence for ComK-binding (Hamoen *et al.*, 1998), but also demonstrated that a large degree of natural variation is allowed in ComK-regulated boxes. The most striking characteristic of K-boxes is the natural variation in spacing length, consisting of two, three or four helical turns in between the start positions of the two AT-boxes of a K-box. Furthermore, K-boxes with up to three deviations from the consensus sequence have been shown to be activated by ComK, resulting in gene transcription. In the present study, an attempt was made to elucidate the role of the different elements of a *B. subtilis* K-box in regulation by ComK, which could explain why at least part of the 92% unregulated K-boxes in the *B. subtilis* genome are not used by ComK to activate gene transcription.
It was demonstrated that the difference in length of the spacing between the AT-boxes, as seen in the three classes of K-boxes, influences the levels of transcription activation and DNA-binding by ComK (fig. 2). Correlating with results of DNA-arrays (Hamoen *et al*., 2002), it is shown that ComK activates transcription at a class III K-box to a lower level than at class I and II K-boxes, likely because binding to class III K-boxes requires higher ComK-concentrations to achieve similar shifts as for the other classes (fig. 2B). A natural example of a class III K-box is the *comK*-promoter. Of this box, it is known that efficient regulation by ComK requires binding of DegU, to prime ComK-binding and transcription activation at the onset of competence development, when ComK-concentrations in the cell are low (Hamoen *et al*., 2000). It

might be that additional regulators are involved also in other class III K-boxes to stimulate transcription activation by ComK.
Previous research demonstrated that binding of ComK results in DNA-bending of the promoter region, as was shown at the *comG*- and *comF*-promoters, where bending of 60-70º was determined (Hamoen *et al*., 1998). DNA-bending is known to be easier for the more flexible AT-rich sequences than for GC's (Koo *et al*., 1986; Perez-Martin *et al*., 1994), but, surprisingly, increasing the GC-content of the spacing from <40 to 60% did not significantly influence regulation by ComK, nor did the positioning of GC-basepairs throughout the spacing region. Apparently, the sequence of the spacing is not an important determinant for the level of regulation by ComK. In contrast to the spacing, the sequence of the AT-box repeats is more critical for determining whether or not a K-box has a high chance of being ComK-regulated as was determined by introduction of A-G or T-G mutations in one AT-box in a further idealized K-box background. In the set-up of this experiment, it should be noted that substitution of a T by a G might have a larger effect than substitution of an A by a G, since in the latter situation the basepair remains a purine, while mutation from a T to a G also implies a change from a pyrimidine to a purine. Indeed, our study reveals larger effects on T-stretches than on A-stretches (fig. 4), so it should be noted that the latter effects might be an underestimate of the importance of the positions in the A-stretch.
Using the single-basepair mutational scanning of a K-box, it was shown that, although large natural variations are allowed, a single substitution of the second thymine residue in AT-box 1 by guanine decreased transcription activation about 3-fold compared to a perfect K-box (fig. 4). The reduction in transcription activation could be explained entirely by a decrease in DNA-binding, which was reduced to a similar level of 30% of the control K-box level. This is however not the case for all positions in the consensus sequence of a K-box, like for example in the case of a T_3- or A_3-mutation in AT-box 2 (fig. 4). In both cases, binding is more affected than transcription activation. It might be that the level of ComK-binding is reduced, but that it is still enough to drive transcription activation to the same level as the control K-box, or, for A_3-G to an even higher level. However, for most positions in the AT-box, binding and transcription activation by ComK are affected to similar extents.

Since a single mutation at the T_2-position of a K-box was shown to most strongly reduce the activity of this box in regulation by ComK, additional experiments were performed to investigate the importance of the T_2-positions. Transcription activation and DNA-binding assays demonstrated that the introduction of a second T_2-mutation in the K-box, reduced transcription and DNA-binding to below 5% of control levels, suggesting that a combination of mutations at both T_2's is not allowed in ComK-regulated boxes (fig. 6). This result was corroborated by *in silico* analyses, demonstrating the absence of double T_2 mutations among ComK-regulated genes (fig. 7). In contrast, for all other individual positions in an AT-box, mutations at the same positions in both boxes could be found among ComK-activated K-boxes. In addition to the T_2-position in AT-box 1, also the basepairs at positions A_2 and A_4 in AT-box 2 are conserved in the case of a T_2-mutation in AT-box 2 (fig. 7B). Whether this is of biological importance or just due to the relatively low number of *in vivo* occuring T_2-mutated boxes among ComK-activated K-boxes, cannot be concluded from this study.
A possibility to maintain activity on a T_2-mutated K-box could be the introduction of mutations in other basepairs in the AT-boxes. Often, protein binding sites display direct or inverted symmetry in the recognition site, like it was demonstrated for the *E. coli* cyclic AMP receptor protein (CRP), the arginine repressor ArgR and the *lac*-repressor LacI (O'Neill *et al*., 1981; Charlier *et al*., 1992; Sadler *et al*., 1983). In the case of the K-box, the two AT-boxes form direct repeats, separated by a spacing of two, three or four helical turns. It might be expected that in T_2-mutated K-boxes, a compensating mutation is introduced to maintain activity of the K-box. Concerning the symmetry of the K-box, a possible compensating mutation could be the T_2 in the other AT-box, which indeed was shown to be very important for maintaining K-box activity, once the T_2-position in AT-box 1 is mutated. Each AT-box consists of an inverted repeat formed by the A- and the T-stretch. Concerning this symmetry, another compensating mutation could be the A_3-position in the same box as the T_2-mutation. Sequence comparisons demonstrated that, with the exception of the K-box of the *bdbDC*-operon, all boxes with a T_2-mutation in AT-box 1, contain three deviations of the consensus sequence. However, *in silico* analyses did not clearly point out conserved mutations in other AT-box basepairs, potentially introduced to maintain activity of a K-box with a T_2-mutation in AT-box 1, as can be seen in fig. 7B.

The most interesting question with respect to the importance of the T_2-positions for regulation by ComK is what could be the reason for its dominating role. We demonstrated that a substitution of T_2 by a guanine lowers the efficiency of ComK-regulation, by reducing ComK-binding to the DNA. The shift-type is not changed, so alteration of the T_2-position does not affect for example oligomerization of ComK on the DNA. Although the mechanism behind the reduced DNA-binding by ComK at a T_2-mutated K-box is unclear, it is likely that the T_2-position is the residue that is in closest contact with or has the highest affinity for ComK in protein-DNA interactions. It is tempting to speculate that the positioning of the K-box along the DNA-helix does not allow a guanine at this position. It might be that a GC-basepair affects the minor groove of the DNA-helix, thereby decreasing ComK-binding efficiency at the K-box. This hypothesis is supported by research on the ARG-boxes bound by the arginine repressor protein in *E. coli*. Wang *et al*. (1998) demonstrated that this repressor contacts the DNA at four major and two minor grooves of a long consensus sequence (ARG-box). The minor groove in the centre of the ARG-boxes is facing the repressor protein and was shown to contain almost exclusively AT-basepairs, suggesting that GC-combinations are not allowed in this region. They suggest that these basepairs would introduce an inhibitory group into the narrow minor groove, which could inhibit binding of the Arg-repressor. Furthermore, they report a growing evidence that the methyl groups of thymine-containing basepairs are often essential for protein-DNA interactions. Although the present study describes a first, general overview of the important basepairs in a K-box, the three-dimensional structure of the protein-DNA complex should preferably be elucidated to unravel the exact nature of ComK-DNA interactions at the site of a K-box.

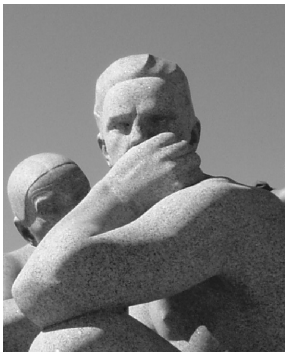
Acknowledgements

We would like to thank Fabrizia Fusetti, Andy-Mark Thunnissen and Wiep Klaas Smits for helpful discussions and Aldert Zomer for help with the *in silico* analyses.

Chapter 5

Expression of a transcription activator, ComK of *Bacillus subtilis*, leads to a genome-wide repression pattern in the heterologous host *Lactococcus lactis*

Kim A. Susanna, Chris D. den Hengst, Leendert W. Hamoen and Oscar P. Kuipers



Abstract

In *Bacillus subtilis* the competence transcription factor ComK plays a central role in the development of genetic competence by activating the K-regulon. ComK-activated genes are characterized by the presence of a ComK-binding site, a K-box, in their upstream region. Each K-box consists of two AT-rich elements (AT-boxes), which are separated by a spacing of a discrete number of helical turns. K-boxes can be found in genomes of other bacteria as well. An interesting example is *Lactococcus lactis*, which contains almost 400 K-boxes located within 200 basepairs from the start of a gene. Despite the presence of K-boxes and a relatively high number of homologues of late competence genes, a *comK*-homologue could not be identified in *L. lactis*. In this study, the effects of production of the transcription activator protein ComK of *B. subtilis* on the transcription profile of the potentially highly responsive host *L. lactis* were investigated using DNA-microarray analyses. Production of wildtype ComK was shown to stimulate the transcription of 89 genes and lower the expression of 114 genes. Notably, potential direct effects, *i. e.* genes preceded by a K-box, were mainly found among the repressed genes, suggesting that ComK here functions as a repressor. This is a remarkable difference with *B. subtilis*, in which ComK almost exclusively acts as an activator. Additional DNA-microarray analyses with a transcription activation-deficient, but DNA-binding ComK-variant, ComKΔC25, demonstrated similar effects on gene regulation as with wildtype ComK, confirming that the direct effects of ComK result from interference with normal transcription through random binding to available K-boxes. This study demonstrates that horizontal gene transfer can have dramatic and very different effects than expected on basis of the original functionality of a gene.

Introduction

The Gram-positive soil bacterium *Bacillus subtilis* is well-known for its abilities to adapt to changes in growth conditions. One of these adaptation processes is genetic competence, a differentiation state that enables the cell to take up DNA from the environment and to incorporate this into its genome. The competent state of a cell is characterized by drastic physiological changes, like switching off DNA replication, cell-wall synthesis and cell division (Haijema *et al*., 2001). Furthermore, a complex DNA-binding, -uptake and -integration machinery is synthesized, which makes the cell competent for transformation.
Competence development is tightly regulated and depends entirely on the presence of the competence transcription factor ComK. Only when the concentration of free ComK in the cell increases, at the onset of the stationary growth-phase, competence can develop (reviews: Dubnau and Lovett, 2002; Hamoen *et al*., 2003). ComK activates gene transcription by binding to ComK-binding sites, so-called K-boxes, which are located upstream of ComK-regulated genes (Hamoen *et al*., 1998). Each K-box consists of two AT-boxes (consensus: AAAA-$(N)_5$-TTTT), separated by a spacing, which positions both AT-boxes on the same side of the DNA-helix (fig. 1). The spacing spans an interval of two, three or four helical turns calculated between the first A's of both AT-boxes.

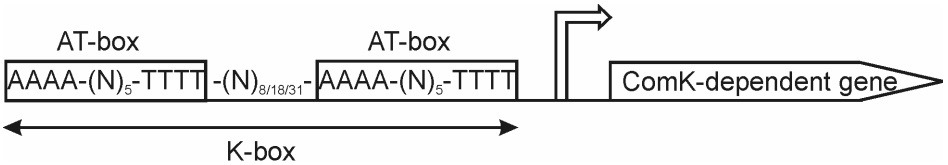


Figure 1. Overview of a K-box
ComK-regulated genes are characterized by the presence of a K-box, upstream of their promoter (arrow). Each K-box consists of two AT-boxes, which are separated by a flexible spacing. This results in positioning of both AT-boxes on the same side of the DNA-helix with an interval of two, three or four helical turns, depending on the length of the spacing, between the start of the two AT-boxes.

Previous research demonstrated that K-boxes with up to three deviations from the consensus sequence can still be good targets for transcription activation by ComK (Hamoen *et al*., 1998). However, not all variations from the consensus sequence are allowed

as could be concluded from a box-scanning mutational analysis, which showed the crucial importance of the second thymine in each AT-box (chapter 4). Furthermore, transcriptome analyses demonstrated that only 8% of the putative functional K-boxes in the *B. subtilis* genome were activated by ComK (Hamoen *et al*., 2002). The inactivity of the other 92% does not necessarily result from deviations from the consensus sequence, but was shown to be caused by for example a wrong positioning relative to the start of a gene (Hamoen *et al*., 2002). Nevertheless, the presence of over a 1000 putative target sites for ComK-binding, of which 30% is located in intergenic regions, reflects a large potential for ComK to act on transcriptional regulation. Searching genome sequences revealed putative K-boxes in other bacteria as well, ranging from small numbers in GC-rich bacteria to large numbers in AT-rich species. *Lactococcus lactis*, for example, contains around 1350 putative K-boxes, of which also roughly one-third (almost 400 K-boxes) is located in promoter regions. Despite the presence of K-boxes as well as homologues of some of the known competence genes, *L. lactis* does not contain *comK* itself, suggesting that the K-boxes are non-functional. It is intriguing to speculate on what would happen to the transcription profile of *L. lactis* upon introduction of ComK, which could be considered as a case study for horizontal gene transfer. Previous research showed that ComK of *B. subtilis* is active when produced in *L. lactis*, as was demonstrated by activation of a *B. subtilis comG*-promoter (chapter 3). It might therefore be expected that ComK, as a putative pleiotropic regulator, interferes extensively with the transcription profile of the host by direct binding to a large number of K-boxes. To investigate the effects of the introduction of ComK on the transcription profile of *L. lactis*, DNA-microarray studies were performed comparing a wildtype *L lactis* strain and two ComK-producing strains, expressing either wildtype ComK or ComKΔC25, a transcription activation-deficient, but DNA-binding mutant. Both ComK-variants affected the transcription profile of about 200 genes, of which the majority was downregulated. The frequency of occurence of a K-box within 200 bp from the start of a gene was higher for downregulated than for upregulated genes, suggesting that ComK excerts a repressive effect on gene transcription in *L. lactis*, mainly because of binding to available K-boxes. However, the majority of the regulated genes were indirectly affected as a consequence of ComK-production in the cell.

Materials and methods

Bacterial strains, media and growth conditions
The bacterial strains that were used in this study are listed in table 1. *L. lactis* strains were grown at 30ºC in two-fold diluted M17-based medium (Difco), supplemented with 0.5% glucose (GM17) and, if required, chloramphenicol and/or erythromycin (both 4 µg/µl). ComK-expression was achieved by induction of the nisin-inducible promoter with 1:1000 dilutions of supernatant of an overnightculture of the nisin-producing strain *L. lactis* NZ9700, which was grown in GM17-medium (de Ruyter *et al*., 1996).

Table 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Relevant feature	Reference
Strains		
L. lactis NZ9000	MG1363 *pepN::nisRK*	Kuipers *et al*., 1998
L. lactis NZ9700	Nisin producing, containing Tn5276	Kuipers *et al.*, 1993
Plasmids		
pNZ8048	Cm^r, inducible expression vector containing the *nisA* promoter	Kuipers *et al*., 1998
pNZ-His_6-ComK	Cm^r, *his_6-comK* under control of P_{nisA}	Chapter 3
pNZ-ComKΔC25	Cm^r, *his_6-comKΔC25* under control of P_{nisA}	Chapter 3
pG-wt	Ery^r, contains wildtype common K-box of *comG*-promoter	Chapter 4
pPywcC	Ery^r, contains P_{ywcC}-*lacZ* fusion	This study

DNA manipulations, materials and transformations
Standard molecular biological techniques were used as described (Ausubel *et al*., 1998; Sambrook *et al*., 1989). Enzymes were purchased from Roche, New England Biolabs or Pharmacia. Plasmids were isolated using the High Pure Plasmid Isolation Kit and PCR products were purified with the High Pure PCR Product Purification Kit (both from Roche). *L. lactis* strains were transformed by electrotransformations using a gene-pulser (BioRad Laboratories) as described by Leenhouts and Venema (1993).

PCR amplifications and plasmid constructions
PCR reactions were performed as described by Innes and Gelfand (1990), using Pwo or Expand DNA-polymerase (both from Roche) on chromosomal DNA of *B. subtilis* 168 or *L. lactis* MG1363 as a template. Plasmids used in this study are listed in table 1. To confirm the effects demonstrated by DNA-microarray analyses on transcription of *L. lactis* genes, around 200 bp of the promoter region of a model-gene was fused

to a promoterless *lacZ*-gene in plasmid pILORI4. For this gene, *ywcC*, the promoter region was amplified using primers ywcC-start (5'-GATCGA ATTCGAAAGCTATCCTACCCCCCTTTC) and ywcC-end (5'-GATCTCTAGATT AAGATACACGTTTAGTATAACCGCC). In both primers, the annealing sequence is underlined. The resulting PCR product was digested with *Xba*I and *Eco*RI and ligated into *Xba*I/*Eco*RI-digested pILORI4, yielding plasmid pPywcC.

ComK-production and transcription activation assays
B. subtilis ComK was produced in *L. lactis* NZ9000, a MG1363 derivative, using plasmid pNZ-His_6-ComK for wildtype ComK (wtComK) or plasmid pNZ-His_6-ComKΔC25 for a C-terminal truncation variant of ComK. These plasmids contain the *comK*-gene under control of the nisin-inducible promoter and induction with nisin results in the synthesis of wtComK or ComKΔC25, respectively. Both ComK-proteins are synthesized with an N-terminal his-tag fusion.
To determine the activity of the produced ComK-proteins, transcription activation assays were performed using the *comG*-promoter-*lacZ* fusion on plasmid pG-wt as a reporter. For this purpose, ComK-expression was induced by adding nisin-containing supernatant after three hours of growth. Samples for β-galactosidase assays were taken from the moment of induction until two hours after induction, with 30 minutes intervals. β-galactosidase activity was determined as described by Israelsen *et al.* (1995). The same experiments were performed using the promoter-*lacZ* fusion on plasmid pPywcC as reporter. ComK-production levels were checked using SDS-PAGE (Laemmli *et* al., 1970) and Western blot analysis (Towbin *et al*., 1979), using a his-tag-specific first antibody and an anti-rabbit horse-radish peroxidase secondary antibody (both from Amersham). The signals were visualized by chemiluminescent detection using the ECL Western blotting analysis system from Amersham.

RNA isolation, cDNA labeling and hybridization
To determine the effect of wtComK-production on the transcription profile of *L. lactis*, cultures of *L. lactis* NZ9000 with plasmid pNZ8048 and with plasmid pNZ-His_6-ComK were grown for three hours until the end of the exponential growth-phase and were then induced with 1:1000 supernatant of the nisin-producing strain *L. lactis* NZ9700. After two hours of induction, cells from 25 ml of culture were harvested by centrifugation (1 min., 8,000g, eppendorf centrifuge) and frozen in liquid nitrogen. Three biological replicates were performed under identical conditions. mRNA was isolated from the cells of each culture as described before

(van Hijum *et al*, manuscript in preparation). cDNA was obtained by reverse transcription and labeled with Cy3 or Cy5. As a control, dye swap reactions were performed under the same conditions. The labeled cDNA samples were hybridized onto an *L. lactis* MG1363 microarray slide as described before (Kuipers *et al.*, 2002; van Hijum *et al*., manuscript in preparation). To determine the effect of ComKΔC25 on the transcription profile of *L. lactis*, the above described procedure was repeated with cultures of *L. lactis* NZ9000 with plasmid pNZ8048 and with plasmid pNZ-His_6-ComKΔC25.

Bioinformatic analyses

After scanning the signals on the slides, bad spots and spots with too low intensities were removed from the data sets using Array Pro analyzer 4.5 (MediaCybernetics, Gleichen, Germany). Based on the assumption that expression of most of the genes is not changed between the two compared situations, the Cy3/Cy5 ratios were normalized using a grid-based Lowess fit (Workman *et al*., 2002). Further analysis was performed using the MicroPrep software package, with subsequent PrePrep, Prep and PostPrep analysis (Van Hijum *et al.*, 2003; Garcia de la Nava *et al*., 2003) and a *t*-test (Hatfield *et al*., 2003). The MicroPrep software package can be requested at: http://molgen.biol.rug.nl/molgen/research/molgen software.php.

***L. lactis* MG1363 genome sequence**

The genome of *L. lactis* MG1363 has been sequenced in a consortium of the University College of Cork in Ireland, the Institute for Food Research in Norwich, United Kingdom and the Department of Molecular Genetics at the University of Groningen, The Netherlands. This resulted in the annotation of 2862 putative open reading frames (Zomer *et al*., manuscript in preparation).

Identification of functional categories among regulated genes

Functional categories among the regulated genes were identified using the Functional Information Viewer and Analyzer (FIVA) program (Blom *et al*., manuscript in preparation).

Genomic overview picture and K-box positioning

The genome2D software package (Baerends *et al*., 2004) was used to create an overview picture of the position of regulated genes on the genome of *L. lactis* MG1363. Furthermore, the program was used to indicate the position of the nearest K-box for each gene.

Results

B. subtilis *ComK is active in* L. lactis*, but affects culture-growth*
The competence transcription factor ComK activates gene transcription in *B. subtilis* by binding to K-boxes in the upstream region of ComK-dependent genes. Like *B. subtilis*, the genome of *L. lactis* contains a large number of K-boxes, which could be potential targets for transcription regulation by ComK. However, *L. lactis* lacks a *comK*-gene itself. Chapter 3 demonstrated the activity of *B. subtilis* ComK in *L. lactis* by showing transcription activation at a *B. subtilis comG*-promoter-*lacZ* fusion. In principle, it might be expected that ComK of *B. subtilis* also regulates transcription from the available K-boxes, when expressed in *L. lactis*. To test this hypothesis, DNA-microarray analyses were performed, comparing the transcription profiles of *L. lactis* NZ9000 with plasmid pNZ8048 and plasmid pNZ-His_6-ComK, both grown in the presence of nisin-containing supernatant to induce transcription from the nisin-inducible promoter on the plasmids. In order to obtain significant array-data, three biological replicates of each strain were produced. Until the moment of induction, after three hours of growth, all cultures displayed a similar growth curve. However, upon induction of ComK-production the growth-rates for the three cultures of *L. lactis* NZ9000 + pNZ-His_6-ComK were reduced, indicating that *comK*-expression affects growth of the host cell (results not shown). Furthermore, Western blots were performed, demonstrating that ComK was indeed produced, at similar levels as in chapter 3 (results not shown)

ComK-production affects 7% of the L. lactis *genome*
To determine the effect of the introduction of *B. subtilis* ComK on the transcription profile of the *L. lactis* host cell, a transcriptome analysis was performed. Including dye-swaps of each of the three biological replicates, six slides were used with a comparison between wildtype *L. lactis* and a ComK-producing strain. The transcription profile of the latter strain revealed a significant effect on about 200 genes, corresponding with 7% of the genome. Notably, 89 genes (3.0% of the genome) were upregulated, while 114 genes (4.0% of the genome) were less expressed in the presence of ComK. The regulated genes with the highest fold differences are listed in table 2 and 3, for up- and downregulated genes respectively.

Table 2. Top 30 of upregulated genes in wtComK or ComKΔC25 expression
The list of genes is based on the result of a wtComK-producing strain. For each gene, the putative function as well as the ratio of transcription relative to *L. lactis* NZ9000 + pNZ8048 is indicated for wtComK and, if regulated, also for ComKΔC25. The last column indicates the presence of a K-box within 200 bp from the transcription start site.

Gene	Description	wtComK	ΔC25	K-box
dacA	D-alanyl-D-alanine-carboxypeptidase	9,8	5,5	
orf1837	Hypothetical protein	8,8	11,2	
orf258	Hypothetical protein	8,2	13,0	
yriC	Hypothetical protein	6,5	15,9	yes
ydaG	ABC-transporter ATP-binding and permease protein	6,3	5,7	
dnaG	DNA primase	6,2	3,9	yes
yjaE	Hypothetical protein	6,2	5,4	
guaC	GMP reductase	5,0	2,4	yes
cadA	Cadmium efflux ATPase	4,9	3,7	
yrbB	Hypothetical protein	4,7	2,2	
yhfA	Hypothetical protein	4,7	2,6	
ymfD	Integrase-recombinase	4,7	4,2	
yriD	Hypothetical protein	4,6	4,7	yes
adhE	Alcohol-acetaldehyde dehydrogenase	4,5	6,3	
ydbA	ABC-transporter ATP-binding and permease protein	4,3	4,7	
yleF	Transcription regulator	4,1	2,2	
ycgD	Oxidoreductase	4,0	3,5	
vacB1	Ribonuclease	4,0	3,5	
trxH	Thioredoxin H-type	3,8	2,0	
orf1734	Hypothetical protein	3,7		
ytjD	Hypothetical protein	3,7	6,0	
yuiE	Hypothetical protein	3,6		
busAA	OpuAA	3,4	3,5	yes
ywjF	3-hydroxyisobutyrate dehydrogenase	3,4		
purM	Phosphoribosylformylglycinamide cyclo-ligase	3,2		
yccL	Hypothetical protein	3,1		
ypaD	Hypothetical protein	3,1	6,6	
yejC	Hypothetical protein	3,1		
yxdB	Hypothetical protein	3,1	2,4	yes
ycdG	Hypothetical protein	3,1	16,0	

Table 3. Top 30 of downregulated genes in wtComK or ComKΔC25 expression
Putative functions, fold regulation and presence of a K-box are indicated as in table 2.

Gene	Description	wtComK	ΔC25	K-box
agl	Alpha-glucosidase	-15,1	-22,2	
mapA	Maltose-phosphorylase	-13,4	-7,8	
ywcC	Hypothetical protein	-11,4	-5,1	yes
pepC	Aminopeptidase C	-11,1		
dexC	Neopullulanase	-11,0	-7,7	
malE	Hypothetical protein	-7,9	-7,0	
tig	Trigger factor	-7,7	-2,0	yes
amyY	Alpha amylase	-6,1	-7,5	
ytaA	Hypothetical protein	-5,4	-15,6	
ydbF	Transcription regulator	-5,2	-7,5	
yjG	Hypothetical protein	-5,2	-3,5	yes

Table 3, continued

Gene	Description	wtComK	ΔC25	K-box
cydA	Cytochrome D ubiquinol oxidase subunit I	-5,1	-2,3	
pgmB	Metabolic protein	-4,6	-5,5	yes
ydbH	Hypothetical protein	-4,5	-8,6	
dnaN	DNA polymerase III, beta-chain	-4,5	-2,0	
yjaB	Oxidoreductase	-4,2	-3,2	
orf1188	Hypothetical protein	-4,2	-4,1	
osmC	Osmotically induced protein	-4,2	-5,0	
maa	Maltose O-acetyltransferase	-4,1	-4,2	
yvdD	Hypothetical protein	-4,0	-5,2	
ffh	Signal recognition particle protein	-4,0	-3,3	
nusB	Transcription termination protein	-4,0		
pepN	Aminopeptidase N	-3,8	-4,8	
malG	Hypothetical protein	-3,8	-3,0	
dexA	Oligo-1,6-glucosidase	-3,8	-3,8	
amtB	NgrA-like protein	-3,7	-2,9	
yseI	Hypothetical protein	-3,6	-3,0	
yjbD	Hypothetical protein	-3,6	-3,0	
pdhD	Lipoamide dehydrogenase comp. of PDH complex	-3,5		
accA	acetylCoA-carboxylase carboxyl transferase α-sub.	-3,5	-2,6	

Genomic organization of ComK-affected genes

The positioning of the regulated genes on the genome may provide information about organization in operons or as individual genes. The genomic organization of the ComK-affected genes was visualized using the Genome2D software package. Expression of ComK was shown to influence transcription of mainly individual or paired genes, but some larger operons could also be discerned. The most eye-catching operon consists of *dexCA*, *maa*, *amyY*, *agl* and *mapA*, which belong to the top-downregulated genes. The gene products are likely involved in sugar metabolism of the cell. Downregulation of this operon can therefore probably be explained by the reduced growth-rate of the ComK-producing strain upon induction of ComK-expression. Furthermore, some members of the *opp*-operon are downregulated, probably also because of the effects of ComK-expression on growth-rate of *L. lactis*. Other clear downregulated operons are the *ycgFGHIJ* operon, the *rnc*, *smc, yibCD* operon and the *ndrHIE* operon. Only two relatively small upregulated operons could be observed: *ydaFG, ydbA* and *ydjB, trxH, ydjD*.

Most of the ComK-effects on the transcription profile are indirect

An interesting question concerning the effects of production of ComK on the transcription profile is whether these effects are direct or indirect. Direct effects are considered to be caused by binding of ComK to a ComK-binding site, *i. e.* a K-box, upstream of a

gene. Previous research in *B. subtilis* on the predictive value of the presence of a K-box for regulation by ComK revealed that K-boxes with 13 or more matches with the 16 basepairs consensus sequence and located within 200 basepairs upstream of the start of a gene could be good targets for regulation by ComK (Hamoen *et al*., 2002). In the present study, a list of K-boxes upstream of the ComK-affected genes was generated, taking into account the important characteristics determined in *B. subtilis* (tables 2 and 3). When this list is considered, it can be concluded that the majority of the effects of ComK-expression is indirect, since only 11 out of the 89 upregulated genes (12,3%) and 31 out of the 114 downregulated genes (27,2%) contain a K-box. Transcriptome analyses in *B. subtilis* indicated that there the frequency of occurrence of a K-box is higher for ComK-activated genes and operons (45%; Hamoen *et al*., 2002), while no significant repression by ComK has been reported (Berka *et al*., 2002; Hamoen *et al*., 2002; Ogura *et al*., 2002).
The low percentage of directly ComK-affected genes, raises questions about the nature of the ComK-induced effects. In order to obtain more information about the background of these effects, functional categories were searched among the regulated genes by the software package FIVA. An overview is shown in table 4. The FIVA-analysis demonstrated that the upregulated genes were mainly involved in protein synthesis. The downregulated genes are grouped into two main categories, *i.e.* ATP-binding and transport and three smaller groups involved in carbohydrate metabolism, regulation of transcription and signal transduction. The observed categories suggest that the induced effects result mainly from the reduced growth in a ComK-producing culture.

Table 4 . Organization of regulated genes in functional categories
Using the FIVA-program, regulated genes have been grouped into functional categories. For each functional category the number of downregulated or upregulated genes in the presence of wtComK is indicated as well as the number of these genes that contain a K-box.

Functional category	Downregulated (K-box)	Upregulated (K-box)
ATP binding	17 (5)	5 (1)
Membrane/transport	20 (3)	11 (1)
Carbohydrate metabolism	7 (0)	3 (0)
Regulation of transcription	7 (2)	2 (0)
Signal transduction	6 (2)	0 (0)
Protein biosynthesis	1 (1)	5 (0)

ComKΔC25 affects transcription to a similar extent as wtComK
The most intriguing observation concerns the fact that most of the direct effects excerted by ComK involve downregulation of gene expression, although ComK in *B. subtilis* almost exclusively acts as a transcription activator (Berka *et al.*, 2002; Hamoen *et al.*, 2002; Ogura *et al*., 2002). This extensive downregulation of gene transcription in *L. lactis* might result from binding of ComK to available K-boxes. Following this reasoning, ComK would disturb the normal transcription regulation by binding to K-boxes. To test this hypothesis, a similar DNA-microarray experiment was performed with a transcription activation-deficient ComK-variant, His_6-ComKΔC25. As described in chapter 3, transcription activation by this mutant is abolished, while DNA-binding is retained with wildtype affinity, despite binding as dimers instead of tetramers. The expression level of ComKΔC25 was checked on Western blots (results not shown), demonstrating a similar protein level as for wildtype ComK (wtComK). Transcriptome analyses demonstrated that production of ComKΔC25 affected the transcription profile of *L. lactis* to a similar extent as wtComK. In total, 237 genes (8,3% of the genome) were affected in the presence of ComKΔC25, divided in 105 upregulated genes (3,7% of the genome) and 132 downregulated genes (4,6% of the genome).
Comparison of the genes affected by wtComK and ComKΔC25 demonstrated that the majority of the top 30 up- and downregulated genes was regulated in both ComK-producing strains (table 2 and 3). Overall, 56, 6% of the wtComK-affected genes were also affected by ComKΔC25. Given the high similarity for the top-regulated genes, this percentage seems relatively low. However, it should be considered that the differences mainly occur at the lower-fold regulated genes, which might just miss the cut-off of fold-regulation or significance in either of the two lists. Furthermore, without exceptions, genes upregulated in a wtComK-expression strain are also upregulated or unchanged by ComKΔC25-production and vice versa, but never downregulated. The same accounts for the downregulated genes. The high similarity in the transcription profiles of wtComK and ComKΔC25 suggests that ComK influences gene regulation by binding to available K-boxes, which interferes with normal gene transcription, thereby explaining the direct down-effects. More often however, ComK-production, which probably causes various stress responses, results in the large number of indirect effects.

β-galactosidase assays show repression by both ComK-variants
To confirm the negative effect of ComK on transcription of *L. lactis* genes, a promoter-*lacZ* fusion was made for a target gene, in this case *ywcC*. This gene was chosen because it is the strongest ComK-repressed gene, with a K-box within 200 bp upstream of the starting position. Transcription of *ywcC* was affected by both wildtype and mutant ComK. A β-galactosidase assay, using plasmid pPywcC as a reporter, demonstrated that transcription from the *ywcC*-promoter is high in the absence of ComK and is reduced when wtComK is produced (fig. 2).

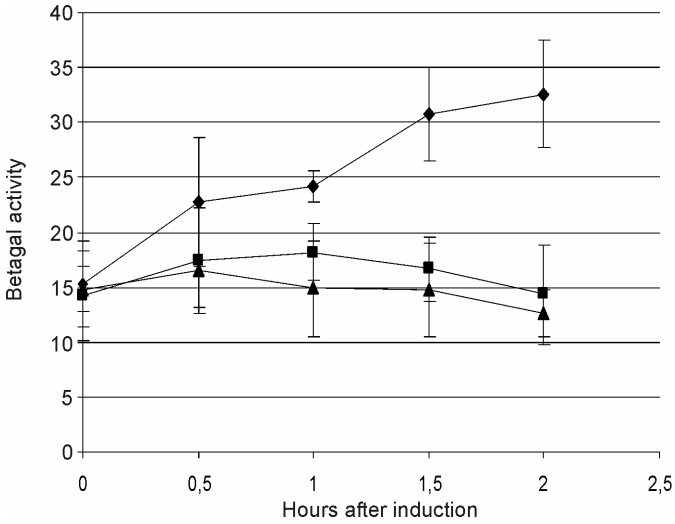


Figure 2. ComK negatively affects transcription of *ywcC*
B-galactosidase assays were performed to demonstrate the negative effect of ComK-expression on transcription of *ywcC*, using a P_{ywcC}-*lacZ* fusion as a reporter. At the end of the exponential growth-phase, expression of wtComK or ComK ΔC25 was induced. B-galactosidase activity was determined with 30 minutes intervals until two hours after induction. For each strain, four independent cultures were grown and the average activity in Miller Units is shown as well as the standard deviation over the average. Diamonds: no ComK (empty plasmid pNZ8048); Triangles: wtComK; Squares: ComKΔC25

A comparable reduction of the transcription level is seen when ComKΔC25 is expressed. Although the ratio between transcription of *ywcC* in wildtype *L. lactis* and in the ComK-expression strains is less than indicated in table 3, these results confirm the trend that was observed using DNA-microarray analyses.

Discussion

The competence transcription factor ComK of *B. subtilis* activates gene transcription of ComK-dependent genes, which are characterized by the presence of a K-box in their upstream DNA (Hamoen *et al*., 1998). Interestingly, other bacterial genomes contain K-boxes as well, especially the genomes of AT-rich bacteria, like for example *L. lactis*. However, despite the large number of K-boxes and the presence of homologues of some of the late competence genes, *L. lactis* does not contain a *comK*-homologue. A previous study demonstrated that ComK of *B. subtilis* retained its transcription activation ability at the *B. subtilis comG*-promoter when introduced into *L. lactis* (chapter 3), although transcription was relatively low compared to the level observed in *B. subtilis*. Originating from this study, the hypothesis was proposed that *B. subtilis* ComK could potentially regulate transcription of *L. lactis* genes preceded by a K-box. DNA-microarray analyses were performed to investigate this possibility, demonstrating that production of ComK affects transcription of about 200 genes in *L. lactis*, which can be divided in 89 upregulated and 114 downregulated genes. Only about 12% of the upregulated and 27% of the downregulated genes contain a K-box in their upstream DNA region, indicating that the majority of ComK-induced effects is indirect. The FIVA-software package was used to group the affected genes into functional categories, confirming that the regulation of these genes mainly results from the effects of ComK-production on the cell. For example, ComK-expression leads to downregulation of genes involved in carbohydrate metabolism, which can be explained by the reduced growth of ComK-producing cultures. Another important category is formed by genes involved in protein synthesis, which are clearly upregulated in a ComK-expressing strain. A possible explanation for this upregulation, could be provided by direct downregulation of other genes by ComK. The products of the latter genes will be depleted from the cell, which might react by enhancing protein synthesis in order to still produce the required proteins. When the lists of regulated genes in *L. lactis* are compared with the list resulting from a comparable analysis in *B. subtilis*, hardly any overlap in regulated genes can be observed, although also in *B. subtilis*, genes involved in metabolic pathways and stress responses are activated upon ComK-synthesis. However, the majority of ComK-activated genes in *B. subtilis* is rela-

ted to competence development, while in *L. lactis* no homologues of known competence genes are affected by ComK-production.
Although most of the effects are indirect, also some direct effects could be discerned. The most striking observation is that direct transcription regulation mainly involves downregulation of gene transcription, concluded from the fact that the frequency of occurrence of a K-box is much higher for downregulated genes (27%) than for upregulated genes (12%). This is a remarkable difference with the situation in *B. subtilis*, since in this species no significant downregulation of gene transcription by ComK has been reported (Berka *et al*., 2002; Hamoen *et al*., 2002; Ogura *et al*., 2002). As an explanation, interference with gene transcription due to binding by ComK to available K-boxes is assumed. This hypothesis was confirmed by an additional DNA-microarray study using a transcription activation-deficient, but DNA-binding ComK-variant, ComKΔC25. The lists of regulated genes show an extensive overlap between the two ComK-producing strains, confirming that the direct effects of ComK on the transcription profile of *L. lactis* are mainly caused by selective positional binding of ComK, thereby disturbing normal transcription activation. The negative effects of both wildtype and mutant ComK on transcription of *L. lactis* genes have furthermore been demonstrated using β-galactosidase assays with a promoter-*lacZ* fusion of *ywcC*, which is the highest affected *L. lactis* gene preceded by a K-box, as a reporter (fig. 2). These experiments confirmed transcriptional repression upon ComK-production, most likely via direct binding to the K-box, thereby obstructing transcription activation.
The similarity in upregulated genes between the expression of wildtype and mutant ComK suggests that these effects are indirect rather than direct, also for the 12% of these genes that do contain a K-box. The occurrence of a K-box upstream of these genes is therefore expected to be a coincidence instead of a biologically relevant feature of the regulated genes. Furthermore, two of the 11 K-boxes upstream of upregulated genes have a double T_2-mutation. In chapter 4, it was demonstrated that this feature results in inactivation of the K-box. In contrast, none of the 31 K-boxes preceding downregulated genes has a double T_2-mutation.
It might be concluded from the described results that ComK only excerts a repressing effect on gene transcription in *L. lactis.* However, despite the fact that direct activation is not observed in the transcription profile of *L. lactis*, it cannot be excluded. Previous

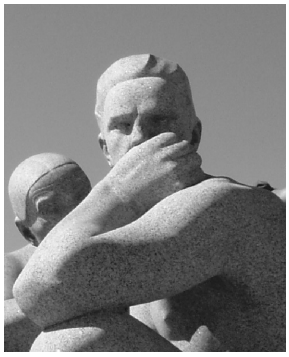
research has demonstrated transcription activation by wtComK at the *comG*-promoter of *B. subtilis* when introduced into *L. lactis*, but the level of transcription was at least 10 times lower than at the same promoter in *B. subtilis*, suggesting that the conditions in *L. lactis* do not allow optimal transcription activation by ComK, which can for example be due to differences in RNA-polymerase. Although it remains speculative, it is likely to expect that ComK of *B. subtilis* does activate transcription by binding to K-boxes in *L. lactis*, but that the transcription level in many cases is too low to be discerned by transcriptome analyses among the large indirect effects, induced by production of ComK.
The most intriguing observation in this study is undoubtedly the change of ComK from a clear activator in *B. subtilis* into a mainly repressing protein in *L. lactis*. The introduction of *B. subtilis* ComK into *L. lactis* could be regarded as an example of horizontal gene transfer between bacteria, demonstrating that a known function of a protein cannot automatically be extrapolated to other potential targets for this protein in another organism. Furthermore, expression of a potentially pleiotropic regulator might induce extensive stress responses and growth-defects in the host organism, thereby interfering with the normal behaviour of the cell. We demonstrated that horizontal gene transfer of a regulatory gene can be an event with highly counterintuitive and unexpected effects compared to its original function.

Acknowledgements

We would like to thank Naomi Kramer for help with performing the DNA-microarray analyses and analysing the obtained data. Furthermore, we thank Andrzej Lulko, Aldert Zomer and Sacha van Hijum for help with the data analysis and Evert-Jan Blom and Anne de Jong for help with the identification of functional categories among the regulated genes.

Chapter 6

Summary and general discussion



Historic overview of competence research

Already in the mid-20th century, natural genetic competence was discovered in *Bacillus subtilis* as an adaptation process, initiated when growth conditions get worse. When a *B. subtilis* population receives environmental signals indicating nutrient deprivation and increased cell densities, part of the culture switches to competence, a differentiation state in which the bacteria are able to take up exogenous DNA from the medium and to incorporate this into their genomes (review: Dubnau and Lovett, 2002).
In the early years of competence research, investigations mainly focussed on the regulatory cascade involved in the development of competence, raising questions as: What are the triggers for initiation? How is competence regulation controlled? What are the output genes and proteins? A major breakthrough was the discovery of the competence transcription factor ComK as the key regulatory protein in competence development (Van Sinderen *et al*., 1995). The regulatory pathway could be divided into several parts, centred around ComK: i) control of ComK-levels in the cell during exponential growth, ii) a system to sense environmental signals on nutritional status and cell densities, and transfer those signals into the cell, where as a response the control of ComK is relieved and iii) activation by ComK of transcription of the late competence genes, encoding the DNA-binding, -uptake and –integration machinery (review: Dubnau and Lovett, 2002; Hamoen *et al*., 2003b). Over the past decade, the picture of the competence regulatory pathway has become more complete by the discovery of novel regulators and by the elucidation of the roles of known regulators involved in the process. A detailed investigation of the DNA-binding and -uptake machinery, led to a better understanding of the DNA-internalization process. Recently, DNA-(micro)-array studies identified additional targets for regulation by ComK, which, together with the already known ComK-regulated genes, were defined as the K-regulon.
Traditionally, in our laboratory the research focus in competence development has been on gene regulation. Previous projects have focussed on the competence regulatory pathway, resulting in the identification of ComK as the central transcription factor (Van Sinderen, 1994), on regulation and function of the competence transcription factor itself (Hamoen, 1998) and on the heterogeneous nature of competence development (Eschevins, 2003).

The work in this thesis can be regarded as a follow-up of the previous research, which identified the consensus sequence of the ComK-binding site as well as the binding characteristics of the protein. Here, we continue on this topic by presenting a functional analysis of the ComK-protein, its target sequences and its mechanism of transcription activation. The present chapter provides a summary of the work described in this thesis as well as an extended discussion of the major results and conclusions. As an overview, the most important results have been summarized and are depicted in a model indicating the requirements for a functional ComK-protein in competence regulation by activating gene transcription in *B. subtilis*. The general discussion concludes by presenting some suggestions for future research.

The mechanism of transcription activation by ComK

ComK activates gene transcription by binding to a K-box located in the upstream region of ComK-activated genes. However, the sole fact that ComK binds to a K-box does not provide an explanation for activation of gene transcription. To investigate the mechanism of transcription activation, the role of ComK in the transcription initiation process was identified, using the promoter of the *comG*-operon as a model. As shown in **chapter 2**, the major effect of ComK on transcription activation is to stabilize RNA-polymerase binding, probably via interactions with the DNA upstream of the K-box, of which the presence was shown to be crucial for transcription activation. These interactions can be facilitated by bending of the DNA in the promoter region, induced by binding of ComK-tetramers. This result was corroborated by the discovery of a tetramerization deficient ComK-mutant, which displays a complete loss of transcription activation (**chapter 3**).
A very intriguing finding in our model system was that the presence of a common K-box was not sufficient for ComK-stimulated gene transcription and stable RNAP-binding *in vitro*. The *comG*-promoter harbours an additional AT-box, located upstream of the two AT-boxes, which form the common K-box, and on the same side of the DNA-helix. The deletion of this extra AT-box resulted in an almost complete loss of transcription activation and stable RNAP-binding *in vitro*. A high level of *in vitro* transcription could only be obtained using a *comG*-promoter, which contains three

AT-boxes. Despite the huge stimulation of *in vitro* gene transcription, the extra AT-box is not required *in vivo*, since most ComK-activated genes contain just a common K-box and show normal transcription levels. It would be interesting to determine the effect on transcription activation of the introduction of an extra AT-box upstream of the common K-box of other ComK-activated genes. Based on the model deduced from the *comG*-promoter, it is expected that introducing an extra AT-box results in measurable *in vitro* transcription at other promoters as well.
Although we first assumed that the presence of an extra AT-box in the *comG*-promoter might be responsible for the high level of transcription activation *in vitro* as well as *in vivo*, more recent studies indicated a discrepancy between the two situations. In *B. subtilis*, transcription at a *comG*-promoter-*lacZ* fusion was higher at a *comG*-promoter with only a common K-box than at a *comG*-promoter with three AT-boxes. However, the results of transcription assays in *L. lactis* correlated with the *in vitro* results of **chapter 2** by demonstrating a higher intrinsic transcription activation ability by *B. subtilis* ComK at a promoter with three AT-boxes. These results suggest additional regulation of *comG*-transcription in *B. subtilis*, which is absent from *L. lactis* as well as from the *in vitro* situation. The extra AT-box could be a target for this additional, yet unidentified regulation. More research is required to investigate this hypothesis. Putative regulatory proteins could be identified by a by-pass mutational analysis, *e. g.* a Tn10 random integration approach. This could be performed using a *comG-lacZ* fusion as a reporter, comparing the wildtype *comG*-promoter with a *comG*-promoter with only a common K-box.

Characteristics of ComK: search for functional domains

After the discovery of the competence transcription factor ComK as the central protein in competence development in *B. subtilis*, research concentrated mainly on regulation. The regulatory role of ComK was identified, including the elucidation of the ComK-binding site, as well as the regulation that a cell exerts to control the ComK-levels before and after competence development. The (near) final result of this research is the scheme of the complex competence regulatory pathway centred around ComK. In this cascade, ComK is involved in different types of interactions, *e. g*.

protein-protein interactions, including ComK-oligomerization, and protein-DNA interactions. This complex behaviour suggests that the ComK-protein contains distinct domains responsible for the diverse functions. In **chapter 3** a search for functional domains in ComK is described, demonstrating that the C-terminal part is required for transcription activation, but not for DNA-binding. However, the truncation variant, ComKΔC25, binds DNA as a dimer instead of a tetramer, suggesting that the C-terminal region of ComK is involved in tetramerization. The loss of tetramerization explains the transcription activation-deficiency of this ComK-variant, since binding of ComK-tetramers is accompanied by DNA-bending of the promoter region (Hamoen *et al*., 1998). As decribed, ComK stimulates transcription activation by stabilizing RNAP-binding, which requires the DNA upstream of the K-box (**chapter 2**). Interactions between this DNA and RNAP can be facilitated by bending of the promoter region, resulting from binding of ComK-tetramers.

Although the protein region required for tetramerization of ComK was identified, the exact mechanism underlying the tetramerization process remains to be elucidated. Common oligomerization mechanisms involve structural interactions between protein domains, like coiled-coil domains, hydrophobic interactions between proteins, formation of hydrogen-bonds or electrostatical interactions. Often, a combination of different interactions drives oligomerization of proteins (Larsen *et al*., 1998). In the case of ComK, *in silico* analyses have been performed to gain insight in the protein. As indicated before, no functional domains could be identified, nor could a secondary structure be predicted. Because of this lack of information, possible structural interactions between protein regions in ComK can not be predicted. However, the probability of different oligomerization mechanisms for ComK can be discussed. The amino acid composition of the C-terminal region of ComK reveals that the last 25 amino acids contain ten hydrophobic and ten charged residues (five positive and five negative). Speculating on the tetramerization mechanism, this distribution suggests hydrophobicity, electrostatical interactions or a combination of both to drive tetramerization of ComK. This hypothesis can be investigated by performing site-directed mutagenesis studies, in which the hydrophobic and charged residues could be mutated and the obtained mutant proteins could be tested for tetramerization and transcription activation.

Another intriguing mutant described in **chapter 3** is the hyper-active N-terminal truncation variant of ComK, ComKΔN9, which is 2 - 3 fold more active relative to wildtype ComK, despite a lower level of protein production in *B. subtilis*. The intrinsic stability of ComKΔN9 is expected to be similar to that of wtComK, because neither of the two ComK-proteins displayed protein instability in *L. lactis*. These observations suggest that the N-terminal part of ComK is involved in regulation of protein expression levels in *B. subtilis*. In **chapter 3**, putative roles for the N-terminal region are discussed, indicating that altered interactions between ComKΔN9 and MecA, which could result in increased degradation by the proteolytic ClpCP-complex, are unlikely. Other options could be a reduction of transcription activation at the *comK*-promoter, which might seem unlikely because of the higher transcription activation activity of ComKΔN9. However, transcription activation was determined at the K-box of the *comG*-promoter, which differs significantly from the one of the *comK*-promoter by the length of the spacing region between the AT-boxes (three versus four helical turns) and by the need for other regulators (none for *comG*, while for *comK* DegU is required at low ComK-levels). The N-terminal region of ComK could be required to enable transcription activation from K-boxes with four helical turns spacing or because of effects of other regulators, like DegU. To test this possibility, transcription could be determined using a *comK-lacZ* fusion instead of a *comG-lacZ* fusion. By performing these experiments in a wild-type and in a Δ*degU* background, the effect of DegU could be determined. Taken together, the results suggest that the presence of the N-terminal extension on ComK is a trade-off between an optimal transcription activation and a thus far unidentified role in ComK regulation, stability or activity.
In addition to the described regions, ComK should contain other domains, responsible for functions like DNA-binding, dimerization and interactions with MecA. Although the research described in **chapter 3** did not identify these functional domains, some speculations could be considered. Concerning DNA-binding, the significant reduction in binding affinity of ComKΔC35 relative to wt-ComK and ComKΔC25 suggests that the truncation of ten extra amino acids touches the DNA-binding domain of ComK. This hypothesis is supported by the limited homology of ComK with the DNA-binding domain of the human protein hSRY. Interestingly, like ComK, hSRY contacts DNA via the minor groove and induces

a comparable DNA-bending of the promoter region (Werner *et al*., 1995). The homologous region of ComK and hSRY covers the sequence from 30 to 120 amino acids calculated from the C-terminus of ComK. If this ComK-region would indeed be involved in DNA-binding, it would explain the reduced binding of ComKΔC35. To test this hypothesis, several experiments could be performed, *e.g.* construct larger truncations of ComK and test these for DNA-binding. However, using this method, instable variants might be created, due to the large protein-deletions. Another, perhaps more suitable approach, would be a site-directed mutational analysis, mutating those residues with proven importance for hSRY-DNA interactions. This implies a more specific determination of the DNA-binding domain than the rough truncation approach, but as a consequence, more mutants are probably required to determine the region of ComK-DNA interactions.
Certainly the best approach would be the crystallization of ComK or preferably ComK-DNA complexes. An attempt was made in collaboration with the Department of Biophysical Chemistry at the University of Groningen (Dr. F. Fusetti and Dr. A.-M. Thunnissen). For this purpose, wildtype ComK was overexpressed as an MBP-fusion in *Escherichia coli* and could be purified to high levels. After cleavage of the MBP-ComK fusion in the presence of DNA, stable ComK-DNA complexes were formed. However, despite numerous crystallization experiments under a wide variety of conditions, diffracting ComK-DNA crystals have not been obtained thus far.
An interesting, although speculative hypothesis would be that the DNA-binding domain of ComK might co-localize with the region interacting with MecA, since these interactions are supposed to be mutually exclusive. In favour of this hypothesis it can be mentioned that ComK interacts with the N-terminal part of MecA (Persuh *et al*., 1999), a region enriched in negatively charged residues. If DNA-binding of ComK would mainly be based on polar interactions with the negatively charged DNA, the same domain in ComK could be involved in interactions with the negative charges in the N-terminal region of MecA. However, disagreeing with this proposed mechanism is the fact that the putative DNA-binding domain of ComK is not extremely positively charged. Moreover, if similar to the DNA-binding of hSRY, polar interactions would not be the favoured mechanism, since DNA-recognition by hSRY has been shown to occur via partial intercalation of a nonpolar side-chain in the minor groove of the DNA (King and Weiss, 1993).

Critical determinants for activity of a K-box

One of the best studied interactions of ComK is binding to DNA. After the discovery of ComK as the central regulator of competence development, the consensus sequence of the ComK-binding site, the so-called K-box, was determined in the upstream DNA of ComK-dependent genes (Hamoen *et al*., 1998). While in **chapter 3** ComK-interactions were investigated by mutating the protein partner of the interaction, **chapter 4** describes a mutational analysis of the DNA of the K-box to which ComK binds. This research originated from the observation that, despite large natural variations in spacing between the AT-boxes and in the consensus sequence of a K-box, apparently not all variations are allowed, since only 8% of all the K-boxes in the *B. subtilis* genome are indeed regulated by ComK (Hamoen *et al*., 2002). Supporting this hypothesis, **chapter 4** describes the identification of the second thymine in the T-stretch of each AT-box as critical for ComK-binding and subsequent transcription activation, which are both reduced 3-fold when the T_2-position in one AT-box is mutated and completely abolished when in both AT-boxes the T_2-positions are substituted by a guanine. No other position in an AT-box appeared so critical for ComK-functionality. This observation suggests that the T_2-position is crucial for ComK-DNA interactions, probably by its specific position in the minor groove of the DNA-helix. The importance of the T_2-position was corroborated by *in silico* analyses, which demonstrated that 11 out of the 88 proven ComK-regulated K-boxes contained a T_2-mutation in AT-box 1, while none of the 88 K-boxes contained a double T_2-mutation. For all other positions, double mutations at similar positions in both AT-boxes can be found. These results clearly demonstrate that a functional K-box should have at least one consensus T_2-position. An interesting speculation would be that the requirement for a consensus T_2 results from the binding characteristics of ComK. Hamoen *et al.* (1998) proposed a strong cooperative binding of ComK-dimers: if one dimer binds to an AT-box, binding of the other dimer almost immediately follows. Binding of a single dimer to a K-box is never observed in unmodified K-boxes and ComK-proteins. In case of a T_2-mutated K-box, this would imply that the first dimer binds to the perfect AT-box and, because of the strong cooperative binding, the other dimer can bind even to an imperfect target site, *e. g.* an AT-box with a T_2-N mutation.

Interestingly, the ComK-regulated K-boxes with a T_2-mutation in one of the AT-boxes show mainly T-A substitutions in K-boxes with higher activities, while T-C or T-G substitutions are only observed among less active K-boxes, suggesting that a T-A basepair at this position is more favourable for K-box activity than a G-C basepair. An explanation could be provided by the nature of the basepairs. A-T basepairs are smaller and more flexible than G-C's (Watson and Crick, 1953). It might be that introduction of a G-C basepair at the T_2-position of an AT-box, results in distortion of the minor groove, thereby reducing ComK-binding. To test this hypothesis, it would be interesting to compare the effects on ComK-binding of T_2-N mutations, with N being either G, C or A.

ComK-introduction into *L. lactis* results in gene repression

ComK of *B. subtilis* retains transcription activation activity upon introduction into *L. lactis*, as was demonstrated in **chapter 3** by activation of a *B. subtilis comG*-promoter-*lacZ* fusion. The *L. lactis* genome sequence contains a large number of K-boxes, which can be potential targets for regulation by ComK. Since *L. lactis* does not have a *comK*-gene itself, these K-boxes are considered to be non-functional. However, introduction of *B. subtilis* ComK might activate transcription at these K-boxes and could provide an interesting case study for horizontal gene transfer by introducing a putative pleiotropic regulator in a potentially highly responsive host bacterium. DNA-microarray analyses demonstrated that production of ComK affects transcription of about 7% (200 genes) of the total genome of *L. lactis*. Notably, the majority (114 genes) was downregulated. Judged by the low frequency of occurrence of K-boxes upstream of regulated genes, most of the effects were indirect. The ComK-induced effects could be clustered into the functional categories of protein synthesis (strongly upregulated) and transport, ATP-binding and carbohydrate metabolism, which were all strongly downregulated in the presence of ComK. As indicated above, direct effects by ComK could only account for a low percentage of regulation (12% of the upregulated and 27% of the downregulated genes). In comparison, 45% of the ComK-activated genes and operons in *B. subtilis* are preceded by a K-box (Hamoen *et al*., 2002), while no significant downregulation has been reported (Berka *et al*., 2002; Hamoen *et al*., 2002; Ogura *et al*.,

2002). The absence of downregulation by ComK in *B. subtilis* makes the overrepresentation of gene repression among the direct effects in *L. lactis* especially intriguing. An explanation for this reciprocal effect could be that the transcriptional regulation results from binding of ComK to the available K-boxes in *L. lactis* leading to interference with normal gene transcription. To test this hypothesis, DNA-microarray analyses were performed using ComKΔC25, a variant which is no longer able to activate gene transcription, but can still bind DNA with a comparable affinity as wtComK (**chapter 3**). A large overlap between the lists of genes affected by wildtype ComK or ComKΔC25 could be observed for both up- and downregulated genes, especially at higher ratio's. This suggests that the direct effects of ComK on the transcription profile of *L. lactis* result from binding by ComK, thereby inhibiting normal transcription. In addition to similar direct effects, production of ComKΔC25 elicits largely the same indirect effects, which confirms that *comK*-expression affects gene transcription in *L. lactis*. The upregulated genes with a K-box in their upstream region are also regulated by both ComK-proteins, suggesting that these effects are indirect and that the presence of a K-box is of less importance.
Although wildtype ComK has previously been shown to activate gene transcription at a *B. subtilis* promoter in *L. lactis*, the DNA-microarray analyses do not reveal clear transcription activation by ComK. However, the level of gene transcription in the test situation in *L. lactis* was at least ten times lower than of the same promoter in *B. subtilis*, indicating that the conditions in *L. lactis* are far from optimal for ComK-activated transcription. Probably, wildtype ComK does stimulate transcription of some of the *L. lactis* genes preceded by a K-box, but this level is likely too low to be observed in a DNA-microarray analysis, since the indirect effects induced by ComK are probably much stronger.

Summary of requirements for functional ComK

The research described in this thesis provides evidence for the importance of distinct elements in the DNA at or near a K-box and in the ComK protein which are required for a functional competence transcription factor in competence regulation. In fig. 1, an overview is depicted, which indicates the different elements that

are necessary for transcription activation by ComK. As was shown in **chapter 4**, a functional K-box requires the presence of a consensus basepair at position T_2 in at least one AT-box and preferably in both boxes. Only then, ComK-dimers can bind to a K-box, each dimer binding to one AT-box. As was described in **chapter 3**, the tetramerization domain of ComK is located in the C-terminal region of the protein.

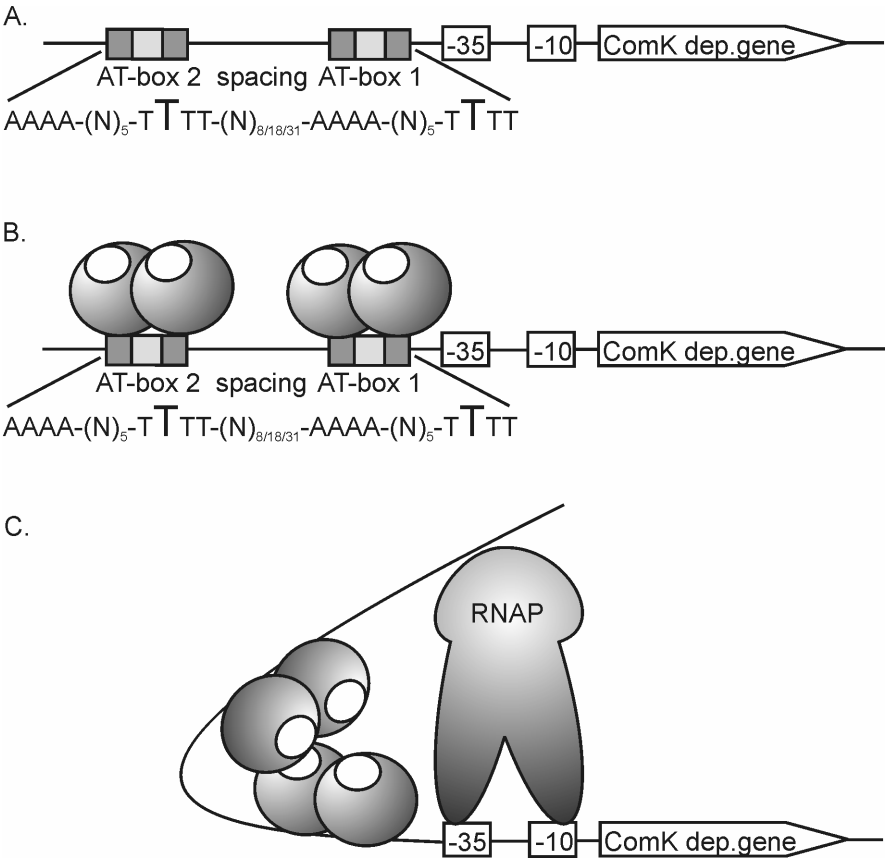


Figure 1. Requirements for a functional ComK in transcription activation
(A). Binding of ComK to a K-box requires a consensus basepair at position T_2 (indicated in capitals) in at least one AT-box. (B). Each ComK-dimer (grey ovals) binds to an AT-box. Tetramerization of ComK requires the C-terminal domain (white) and is accompanied by DNA-bending, shown in panel C. (C). The upstream DNA-region stabilizes RNAP-binding. Interactions between RNAP and the upstream DNA are facilitated by DNA-bending of the promoter region, induced by a ComK-tetramer.

A loss of tetramerization results in a loss of transcription activation by ComK. The research described in **chapter 2** demonstrated that the major role of ComK in transcription activation at the promoter of *comG* is to stimulate binding of RNAP, which requires the upstream DNA. RNAP-binding is proposed to be stabilized through interactions with this upstream DNA, facilitated by DNA-bending of the promoter, which is induced by binding of a ComK-tetramer.

Concluding remarks and suggestions for future research

The research described in this thesis presents a functional analysis of the competence transcription factor ComK of *Bacillus subtilis*. The role of ComK in transcription activation was identified using the promoter of *comG* as a model. Although differences between ComK-dependent promoters, *e. g.* the spacing length or the involvement of other regulators, might require additional effects of ComK, the major role of ComK in transcription initiation at all promoters is considered to be to stabilize RNA-polymerase binding by facilitating interactions with the DNA upstream of the K-box. Coming to this conclusion, the work on the general role of ComK in the transcription activation mechanism can be regarded as finished. However, subsequent projects can of course aim for a detailed investigation of the specific role of ComK in transcription initiation at the *comK*-promoter, which is expected to be more complex because of the longer spacing and the involvement of other regulatory proteins.
The research described in the other three chapters of this thesis, has only just been started and, although interesting results have been obtained, many questions are still unanswered. It would therefore be highly interesting to continue this research in a new project. As a follow-up of **chapter 3**, concerning functional domains in ComK, it would be recommended to focus on the crystal structure of ComK, in solution or, preferably, in interactions with DNA or MecA. Obtaining crystal structures would mean a major breakthrough in ComK-research, since these can help to identify functional domains in ComK, like the DNA-binding domain, the dimerization domain and the site of interaction with MecA, as well as the mechanism driving tetramerization. This approach would also provide a continuation of the research described in **chapter 4**, concerning crucial elements in a K-box for activity of this box,

since the crystal structure of a ComK-DNA complex could help to elucidate the exact role of the T_2-positions in ComK-binding.
Of course, the work described in this thesis is only one of the focuses in competence research throughout the scientific community. An important topic concerns for example the heterogeneous nature of competence development. Recently, two interesting studies about the processes underlying heterogeneity or bistability of ComK expression in *B. subtilis* have been published (Smits *et al*., 2005; Maamar *et al.*, 2005). In addition to regulation of competence, research concentrated on the mechanism of DNA-transport into the cell, resulting in the recent identification of the requirement of the proton motive force for DNA-internalization (Maier *et al*., 2005). Together, these results lead to a better understanding of all processes involved in competence in *B. subtilis*, which can also be extrapolated to other bacterial species in which competence has been identified, resulting in for example the recent identification of ComE as an essential DNA-uptake protein in *Bacillus megaterium* (Lammers *et al*., 2004).

Appendix 1

References

Adhya, S. and S. Garges (1990) Positive Control. *J. Biol. Chem.* 265, 10797-10800

Adhya, S., M. Gottesman, S. Garges, and A. Oppenheim (1993) Promoter resurrection by activators – a minireview. *Gene* 132, 1-6

Aizawa, S.-I., I. B. Zhulin, L. Marquez-Magana and G. W. Ordal (2002) Chemotaxis and motility. In *Bacillus subtilis and its closest relatives. From genes to cells*. Sonenshein, A. L., J. A. Hoch and R. Losick (eds). Washington, D.C., American society for microbiology, pp. 437-452

Albano, M., J. Hahn and D. Dubnau (1987) Expression of competence genes in *Bacillus subtilis*. *J. Bacteriol.* 169, 3110-3117

Anagnostopoulos, C. and J. Spizizen (1961) Requirements for transformation in *Bacillus subtilis*. *J. Bacteriol.* 81, 741-746

Ansaldi, M., D. Marolt, T. Stebe, I. Mandic-Mulec and D. Dubnau (2002) Specific activation of the *Bacillus* quorum-sensing systems by isoprenylated pheromone variants. *Mol. Microbiol.* 44, 1561-1573

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidham, J. A. Smith and K. Struhl (1998) Current Protocols in Molecular Biology. John Wiley and Sons, New York

Baerends, R. J. S., W. K. Smits, A. de Jong, L. W. Hamoen, J. Kok and O. P. Kuipers (2004) Genome2D: a visualization tool for rapid analysis of bacterial transcriptome data. *Genome Biol.* 5, 1-6

Berka, R.M., J. Hahn, M. Albano, I. Draskovic, M. Persuh, X. Cui, A. Sloma, W. Widner and D. Dubnau (2002) Microarray analysis of the *Bacillus subtilis* K-state: genome-wide expression changes dependent on ComK. *Mol. Microbiol.* 43, 1331-1345

Bewley, C. A., A. M. Gronenborn and G. M. Clore (1998) Minor groove-binding architectural proteins: structure, function and DNA recognition. *Annu. Rev. Biophys. Biomol. Struct.* 27, 105-131

Blom, E. J., D. W. J. Bosman, S. A. F. T. van Hijum, J. B. T. M. Roerdink and O. P. Kuipers, manuscript in preparation

Bodmer, W. and A. T. Ganesan (1964) Biochemical and genetic studies of integration and recombination in *Bacillus subtilis* transformation. *Genetics* 50, 717-738

Borowiec, J. A. and J. D. Gralla (1987) All three elements of the *lac* p^S promoter mediate its transcriptional response to DNA supercoiling. *J. Mol. Biol.* 195, 89-97

Boucher, P. E. and S. Stibitz (1995) Synergistic binding of RNA polymerase and BvgA phosphate to the pertussis toxin promoter of *Bordetella pertussis. J. Bacteriol.* 177, 6486-6791

Bracco, L., D. Kotlarz, A. Kolb, S. Diekmann, and H. Buc (1989)

Synthetic curved DNA sequences can act as transcriptional activators in *Escherichia coli. EMBO J.* 8, 4289-4296

Bron, S. and G. Venema (1972) Ultraviolet inactivation and excision repair in *Bacillus subtilis*. I. Construction and characterization of a transformable eightfold auxotrophic strain and two ultraviolet-sensitive derivatives*. Mutat. Res.* 15, 1-10

Brunner, M. and H. Bujard (1987) Promoter recognition and promoter strength in the *Escherichia coli* system. *EMBO J.* 6, 3139-3144

Buc, H (1986) Mechanism of activation of transcription by the complex formed between cyclic-AMP and its receptor en *Escherichia coli. Biochem. Soc. Trans.* 14, 196-199

Burkholder, P. R. and N. H. Giles (1947) Induced biochemical mutations in *Bacillus subtilis*. *Am. J. Botany*, 34, 345-348

Busby, S. and R. H. Ebright (1997) Transcription activation at class II CAP-dependent promoters. *Mol. Microbiol.* 23, 853-859

Carpousis, A. J. and J. D. Gralla (1985) Interaction of RNA polymerase with *lac*UV5 promoter DNA during mRNA initiation and elongation. *J. Mol. Biol.* 183, 165-177

Chang, B. Y. and R. H. Doi (1990) Overproduction, purification, and characterization of Bacillus subtilis RNA polymerase sigma A factor. *J. Bacteriol.* 172, 3257-3263

Charlier, D., M. Roovers, F. van Vliet, A. Boyen, R. Cunin, Y. Nakamura, N. Glansdorff and A. Pierard (1992) Arginine regulon of *Escherichia coli* K-12. A study of repressor-operator interactions and of *in vitro* binding affinities versus *in vivo* repression. *J. Mol. Biol.* 226, 367-386

Chen, I. and D. Dubnau (2003) DNA transport during transformation. *Front. Biosci.* 8, 544-556

Chung, J. D., G. Sephanopoulos, K. Ireton and A. D. Grossman (1994) Gene expression in single cells of *Bacillus subtilis*: Evidence that a threshold mechanism controls the initiation of sporulation. *J. Bacteriol.* 176, 1977-1984

Chung, Y. S. and D. Dubnau (1995) ComC is required for the processing and translocation of ComGC, a pilin-like competence protein of *Bacillus subtilis*. *Mol. Microbiol.* 15, 543-551

Chung, Y. S. and D. Dubnau (1998) All seven *comG* open reading frames are required for DNA binding during transformation of competent *Bacillus subtilis*. *J. Bacteriol.* 180, 41-45

Chung, Y. S., F. Breidt and D. Dubnau (1998) Cell surface localization and processing of the ComG proteins, required for DNA binding during transformation of *Bacillus subtilis*. *Mol. Microbiol.* 29, 905-913

Contente, S. and D. Dubnau (1979) Characterization of plasmid transformation in *Bacillus subtilis*: kinetic properties and the effect of DNA conformation. *Mol. Gen. Genet.* 167, 251-258

Core, L. and M. Perego (2003) TPR_mediated interaction of RapC with ComA inhibits response-regulator-DNA binding for competence development in *Bacillus subtilis*. *Mol. Microbiol*. 49, 1509-1522

Davidoff-Abelson, R. and D. Dubnau (1973) Conditions affecting the isolation from transformed cells of *Bacillus subtilis* of high molecular weight single-stranded deoxyribonucleic acid of donor origin. *J. Bacteriol.* 116, 154-162

De Ruyter, P. G., O. P. Kuipers and W. M. de Vos (1996) Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. *Appl. Environ. Microbiol.* 62, 3662-3667

Déthiollaz, S., P. Eichenberger, and J. Geiselmann (1996) Influence of DNA geometry on transcriptional activation in *Escherichia coli. EMBO J.* 15, 5449-5458

Dooley, D. C., C. T. Hadden and E. W. Nester (1971) Macromolecular synthesis in *Bacillus subtilis* during development of the competent state. *J. Bacteriol.* 108, 668-679

D'Souza, C., M. M. Nakano and P. Zuber (1994) Identification of *comS*, a gene of the *srfA* operon that regulates the establishment of genetic competence in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 91, 9397-9401

Dubnau, D. and C. Cirigliano (1972a) Fate of transforming DNA following uptake by competent *Bacillus subtilis*. III. Formation and properties of products isolated from transformed cells which are derived entirely from donor DNA. *J. Mol. Biol.* 64, 9-29

Dubnau, D. and C. Cirigliano (1972b) Fate of transforming DNA following uptake by competent *Bacillus subtilis*. IV. The end-wise attachment and uptake of transforming DNA. *J. Mol. Biol.* 64, 31-46

Dubnau, D. and M. Roggiani (1990) Growth-medium independent genetic competence mutants of *Bacillus subtilis*. *J. Bacteriol.* 172, 4048-4055

Dubnau, D (1993) Genetic exchange and homologous recombination, p555-584. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington D. C.

Dubnau, D (1999) DNA uptake in bacteria. *Annu. Rev. Microbiol.* 53, 217-244

Dubnau, D. and C. M. J. Lovett (2002) Transformation and recombination. In *Bacillus subtilis and its closest relatives. From genes to*

cells. Sonenshein, A. L., J. A. Hoch and R. Losick (eds). Washington, D. C., American society for microbiology, pp. 453-471

Eichenberger, P., S. Déthiollaz, N. Fujita, A. Ishihama, and J. Geiselmann (1996) Influence of the location of the cAMP receptor protein binding site on the geometry of transcriptional activation complex in *Escherichia coli*. *Biochemistry* 35, 15302-15312

Eichenberger, P., S. Déthiollaz, H. Buc, and J. Geiselmann (1997) Structural kinetics of transcription activation at the *malT* promoter of *Escherichia coli* by UV laser footprinting. *Proc. Natl. Acad. Sci. USA* 94, 9022-9027

Ellinger, T., D. Behnke, H. Bujard, and J.D. Gralla (1994a) Stalling of *Escherichia coli* RNA polymerase in the +6 to +12 region *in vivo* is associated with tight binding to consensus promoter elements. *J. Mol. Biol.* 239, 455-465

Ellinger, T., D. Behnke, R. Knaus, H. Bujard, and J.D. Gralla (1994b) Context-dependent effects of upstream A-tracts: Stimulation or inhibition of *Escherichia coli* promoter function. *J. Mol. Biol.* 239, 466-475

Engelhorn, M. and J. Geiselmann (1998) Maximal transcriptional activation by the IHF protein of *Escherichia coli* depends on optimal DNA bending by the activator. *Mol. Microbiol.* 30, 431-441

Eschevins, C. C. (2003) Regulatory networks related to competence development in *Bacillus subtilis*. Ph. D. Thesis. University of Groningen

Estrem, S. T., T. Gaal, W. Ross, and R. L. Gourse (1998) Identification of an UP element consensus sequence for bacterial promoters. *Proc. Natl. Acad. Sci.* 95, 9761-9766

Estrem, S. T., W. Ross, T. Gaal, Z. W. S. Chen, W. Niu, R. H. Ebright, and R. L. Gourse (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase α subunit. *Genes. Dev.* 13, 2134-2147

Frederick, K., T. Caramori, Y. Chen, A. Galizzi, and J.D. Helmann (1995) Promoter architecture in the flagellar regulon of *Bacillus subtilis*: High-level expression of flagellin by the σ^D RNA polymerase requires an upstream promoter element. *Proc. Natl. Acad. Sci. USA* 92, 2582-2586

Fujita, M., and Y. Sadaie (1998) Rapid isolation of RNA polymerase from sporulating cells of *Bacillus subtilis*. *Gene* 221, 185-190

Garcia de la Nava, J., S. van Hijum and O. Trelles (2003) Prep: gene expression data pre-processing*. Bioinformatics* 19, 2328-2329

Gartenberg, M. R. and D. M. Crothers (1991) Synthetic DNA bending sequences increase the rate of *in vitro* transcription initiation at the *Escherichia coli lac* promoter. *J. Mol. Biol.* 219, 217-230

Gearhart, M. D., S. M. A. Holmbeck, R. M. Evans, H. J. Dyson and P. E. Wright (2003) Monomeric complex of human orphan estrogen related receptor-2 with DNA: a pseudo-dimer interface mediates extended half-site recognition. *J. Mol. Biol.* 327, 819-832

Geierstanger, B. H., B. F. Volkman, W. Kremer and D. E. Wemmer (1994) Short peptide fragments derived from HMG-I/Y proteins bind specifically to the minor groove of DNA. *Biochemistry* 33, 5347-5355

Geiselmann, J (1997) The role of DNA conformation in transcriptional initiation and activation in *Escherichia coli. J. Biol. Chem.* 378, 599-607

Gross, C. A., M. Lonetto, and R. Losick (1992) Bacterial sigma factors, p.129-176. *In* K. Yamamoto and S. McKnight (ed.), Transcriptional regulation. Cold Spring Harbor Laboratory, NY

Gourse, R. L., W. Ross, and T. Gaal (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol. Microbiol.* 37, 687-695

Grosschedl, R., K. Giese and J. Pagel (1994) HMG domain proteins: architectural elements in the assembly of nucleoprotein structures. *Trends Genet*. 10, 94-100

Hadden, C. and E. W. Nester (1968) Purification of competent cells in the *Bacillus subtilis* transformation system. *J. Bacteriol.* 95, 876-885

Hahn, J. L. Kong and D. Dubnau (1994) The regulation of competence transcription factor synthesis constitutes a critical control point in the regulation of competence in *Bacillus subtilis*. *J. Bacteriol.* 176, 5753-5761

Hahn, J., M. Roggiani and D. Dubnau (1995) The major role of Spo0A in genetic competence is to downregulate *abrB*, an essential competence gene. *J. Bacteriol.* 177, 3601-3605

Hahn, J., A. Luttinger and D. Dubnau (1996) Regulatory inputs for the synthesis of ComK, the competence transcription factor of *Bacillus subtilis*. *Mol. Microbiol.* 21, 763-775

Haijema, B. J., L. W. Hamoen, J. Kooistra, G. Venema, and D. Dubnau (1995) Expression of the ATP-dependent deoxyribonuclease of *Bacillus subtilis* is under competence-mediated control. *Mol. Microbiol.* 15, 203-211

Haijema, B. J., D. Van Sinderen, K. Winterling, J. Kooistra, G. Venema, and L. W. Hamoen (1996) Regulated expression of the *dinR* and *recA* genes during competence development and SOS induc-

tion in *Bacillus subtilis*. *Mol. Microbiol.* 22, 75-85

Haijema, B. J., J. Hahn, J. Haynes and D. Dubnau (2001) A ComGA-dependent checkpoint limits growth during the escape from competence. *Mol. Microbiol.* 40, 52-64

Hamoen, L. W., H. Eshuis, J. Jongbloed, G. Venema and D. van Sinderen (1995) A small gene, designated *comS*, located within the coding region of the fourth amino acid-activation domain of *srfA*, is required for competence development in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 97, 9246-9251

Hamoen, L. (1998) The competence transcription factor of *Bacillus subtilis*. Regulation and function. Ph. D. Thesis. University of Groningen

Hamoen, L. W., A. F. van Werkhoven, J. J. E. Bijlsma, D. Dubnau and G. Venema (1998) The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. *Genes Dev.* 12, 1539-1550

Hamoen, L. W., A. F. van Werkhoven, G. Venema and D. Dubnau (2000) The pleiotropic response regulator DegU functions as a priming protein in competence development in *Bacillus subtilis. Proc. Natl. Acad. Sci. U S A* 97, 9246-9251

Hamoen, L.W., W.K. Smits, A. de Jong, S. Holsappel and O.P. Kuipers (2002) Improving the predictive value of the competence transcription factor (ComK) binding site in *Bacillus subtilis* using a genomic approach. *Nucl. Acid. Res.* 30, 5517-5528

Hamoen, L. W., D. Kausche, M. A. Marahiel, D. van Sinderen, G. Venema and P. Serror (2003a) The *Bacillus subtilis* transition state regulator AbrB binds to the –35 promoter region of *comK*. *FEMS Microbiol. Lett.* 218, 299-304

Hamoen, L. W., G. Venema and O. P. Kuipers (2003b) Controlling competence in *Bacillus subtilis*: shared use of regulators. *Microbiology* 149, 9-17

Harwood, C. R. (1992) *Bacillus subtilis* and its relatives: molecular biological and industrial workhorses. *Trends Biotechnol.* 10, 247-256

Haseltine-Cahn, F. and M. S. Fox (1968) Fractionation of transformable bacteria from competent cultures of *Bacillus subtilis* on renografin gradients. *J. Bacteriol.* 95, 867-875

Hatfield, G. W., S. Hung and P. Baldi (2003) Differential analysis of DNA microarray expression data. *Mol. Microbiol.* 47, 871-877

Helmann, J.D. (1995) Compilation and analysis of *Bacillus subtilis* σ^A-dependent promoter sequences: evidence for extended contact be-

tween RNA polymerase and upstream promoter DNA. *Nucl. Acids Res.* 23, 2351-2360

Hilbert, D. W. and P. J. Piggot (2004) Compartimentalization of gene expression during *Bacillus subtilis* spore formation. *Microbiol. Mol. Biol. Rev.* 68, 234-262

Hoa, T. T., P. Tortosa, M. Albano and D. Dubnau (2002) Rok (YkuW) regulates genetic competence in *Bacillus subtilis* by directly repressing *comK*. *Mol. Microbiol.* 9, 365-373

Hoch, J. A. (1971) Genetic analysis of pleiotropic negative sporulation mutants in *Bacillus subtilis*. *J. Bact*, 105, 896-901

Hochschild, A. and S. L. Dove (1998) Protein-protein contacts that activate and repress prokaryotic transcription. *Cell* 92, 597-600

Hsu, L. M. (2002) Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta* 1577, 191-207

Inamine, G. S. and D. Dubnau (1995) ComEA, a *Bacillus subtilis* integral membrane protein required for genetic transformation, is needed for both DNA binding and transport. *J. Bacteriol.* 177, 3045-3051

Innes, M. A. and D. H. Gelfand (1990). Optimization of PCRs, p.3-12. *In* M.A. Innes, D. H. Gelfand, J. J. Sninsky and T. J. White (ed.), *PCR protocols: A guide to methods and applications*. Academic Press, San Diego, CA.

Ishihama, A. (1993) Protein-protein communication within the transcription apparatus. *J. Bacteriol.* 175, 2483-2489

Israelsen, H., S. M. Madsen, A. Vrang, E. B. Hansen and E. Johansen (1995) Cloning and partial characterization of regulated promoters from *Lactococcus lactis* Tn917-lacZ integrants with the new promoter probe vector, pAK80. *Appl. Environ. Microbiol.* 61, 2540-2547

Kanamaru, K., S. Stephenson and M. Perego (2002) Overexpression of the PepF oligopeptidase inhibits sporulation initiation in *Bacillus subtilis*. *J. Bacteriol.* 184, 43-50

King, C.-Y. and M. A. Weiss (1993) The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: A topological mechanism of sequence specificity. *Proc. Natl. Acad. Sci. U.S.A.* 90, 11990-11994

Knaus, R. and H. Bujard (1988) P_L of coliphage lambda: an alternative solution for an efficient promoter. *EMBO J.* 7, 2919-2923

Kobayashi, K, S. D. Ehrlich, Albertini, A., *et al*. (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci.* 100, 4678-4683

Kong, L. and D. Dubnau (1994) Regulation of competence specific

gene expression by Mec-mediated protein-protein interaction in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 91, 5793-5797

Koo, H., H. Wu and D. M. Crothers (1986) DNA bending at adenine-thymine tracts. *Nature* 320, 501-506

Kuipers, O. P., M. M. Beerthuyzen, R. J. Siezen and W. M. de Vos (1993) Characterization of the nisin gene cluster *nisABTCIPR* of *Lactococcus lactis* requirement of expression of the *nisA* and *nisI* genes for development of immunity. *Eur. J. Biochem.* 216, 281-291

Kuipers, O. P., P. G. G. A. de Ruyter, M. Kleerebezem and W. M. de Vos (1998) Quorum-sensing-controlled gene expression in lactic acid bacteria. *J. Biotechnol.* 64, 15-21

Kuipers, O. P., A. de Jong, R. J. S. Baerends, S. A. F. T. van Hijum, A. L. Zomer, H. A. Karsens, C. D. den Hengst, N. E. Kramer, G. Buist and J. Kok (2002) Transcriptome analysis and related databases of *Lactococcus lactis*. *Antonie van Leeuwenhoek* **82,** 113-122

Kunst, F., Ogasawara, N., Moszer, I., *et al.* (1997) The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 390, 249-256

Kunst, F., T. Msadek, J. Bignon and G. Rapoport (1994) The DegS/DegU and ComP/ComA two-component systems are part of a network controlling degradative enzyme synthesis and competence in *B. subtilis*. *Res. Microbiol*. 145, 393-402

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685

Lammers, M., H. Nahrstedt and F. Meinhardt (2004) The *Bacillus megaterium comE* locus encodes a functional DNA uptake protein. *J. Basic Microbiol.* **44,** 451-458

Larsen, R., G. Buist, O. P. Kuipers and J. Kok (2004) ArgR and AhrC are both required for regulation of arginine metabolism in *Lactococcus lactis*. *J. Bacteriol.* 186, 1147-1157

Larsen, T. A., A. J. Olsen and D. S. Goodsell (1998) Morphology of protein-protein interfaces. *Structure* 6, 421-427

Lazazzera, B. A., J. M. Solomon and A. D. Grossman (1997) An exported peptide functions intracellularly to contribute to cell density signaling in *B. subtilis*. *Cell* 89, 917-925

Leenhouts, K. J. and G. Venema (1993) Plasmids, a practical approach. p 65-94. Oxford University Press, Oxford

Levine, J. S. and N. Strauss (1965) Lag period characterizing the entry of transforming deoxyribonucleic acid into *Bacillus subtilis*. *J. Bacteriol.* 89, 281-287

Lewis, D. E. A., M. Geanacopoulos, and S. Adhya (1999) Role of HU

and DNA supercoiling in transcription repression: specializes nucleoprotein repression complex at *gal* promoters in *Escherichia coli. Mol. Microbiol.* 31, 451-461

Londono-Vallejo, J. A. and D. Dubnau (1993) *comF*, a *Bacillus subtilis* late competence locus, encodes a protein similar to ATP-dependent RNA/DNA helicases. *Mol. Microbiol.* 9, 119-131

Londono-Vallejo, J. A. and D. Dubnau (1994a) Membrane assication and role in DNA uptake of the *Bacillus subtilis* PriA analog ComFI. *Mol. Microbiol.* 13, 197-205

Londono-Vallejo, J. A. and D. Dubnau (1994b) Mutation of the putative nucleotide binding site of the *Bacillus subtilis* membrane protein ComFA abolishes the uptake of DNA during transformation. *J. Bacteriol.* 176, 4642-4645

Lorenz, M. G. and W. Wackernagel (1994) Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol. Rev.* 58, 563-602

Lloyd, G., P. Landini, and S. Busby (2001) Activation and repression of transcription initiation in bacteria. *Essays Biochem.* 37, 17-31

Maamar, H. and D. Dubnau (2005) Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Mol. Microbiol.*, in press

Magnuson, R. J. Solomon and A. D. Grossman (1994) Biochemical and genetic characterization of a competence pheromone from *B. subtilis*. *Cell* 77, 207-216

Maier, B., I. Chen, D. Dubnau and M. P. Sheetz (2004) DNA transport into *Bacillus subtilis* requires protein motive force to generate large molecular forces. *Nat. Struct. Mol. Biol.* **11,** 643-649

McClure, W. R. (1985) Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.* 54, 171-204

Meijer, W. J. J., G. Venema, and S. Bron (1995) Alternative, RecA-mediated pathway for the initiation of lagging strand synthesis of the broad-host range rolling-circle plasmid pMV158 in *Bacillus subtilis*. In *Replication and maintenance of plasmids in Bacillus subtilis*. W.J.J. Meijer, Ph.D. thesis, University of Groningen, the Netherlands

Mencia, M., M. Monsalve, F. Rojo, and M. Salas (1996) Transcription activation by phage Φ29 protein p4 is mediated by interaction with the α subunit of *Bacillus subtilis* RNA polymerase. *Proc. Natl. Acad. Sci. USA* 93, 6616-6620

Menendez, M., A. Kolb, and H. Buc (1987) A new target for CRP action at the *malT* promoter. *EMBO J.* 6, 4227-4234

Michod, R. E., M. F. Wojciechowski and M. A. Hoelzer (1988) DNA

repair and the evolution of transformation in the bacterium *Bacillus subtilis*. *Genetics* 118, 31-39

Mishra, R. K., and D. Chatterji (1993) Mechanism of initiation of transcription by *Escherichia coli* RNA polymerase on supercoiled template. *Mol. Microbiol.* 8, 507-515

Mock, M. and A. Fouet (2001) Anthrax. *Annu. Rev. Microbiol.* 55, 647-671

Mohan, S. and D. Dubnau (1990) Transcriptional regulation of *comC*: Evidence for a competence-specific factor in *Bacillus subtilis*. *J. Bacteriol.* 172, 4064-4071

Monsalve, M., M. Mencia, F. Rojo, and M. Salas (1996) Activation and repression of transcription at two different phage Φ29 promoters are mediated by interaction of the same residues of regulatory protein p4 with RNA polymerase. *EMBO J.* 15, 383-391

Msadek, T. F. Kunst and G. Rapoport (1994) MecB of *Bacillus subtilis* is a pleiotropic regulator of the ClpC ATPase family, controlling competence gene expression and survival at high temperature. *Proc. Natl. Acad. Sci. USA* 91, 5788-5792

Msadek, T., V. Dartois, F. Kunst, M. L. Herbaud, F. Denizot and G. Rapoport (1998) ClpP of *Bacillus subtilis* is required for competence development, motility, degradative enzyme synthesis, growth at high temperature and sporulation. *Mol. Microbiol.* 27, 899-914

Msadek, T. (1999) When the going gets tough: survival strategies and environmental signaling networks in *Bacillus subtilis*. *Trends Microbiol.* 7, 201-207

Murakami, K., N. Fujita, and A. Ishihama (1996) Transcription factor recognition surface on the RNA polymerase α subunit is involved in contact with the DNA enhancer element. *EMBO J.* 15, 4358-4367

Nakano, M. M. and P. Zuber (1989) Cloning and characterization of *srfB*, a regulatory gene involved in surfactin production and competence in *Bacillus subtilis*. *J. Bacteriol*. 171, 5347-5353

Nester, E. W. and B. A. D. Stocker (1963) Biosynthetic latency in early stages of deoxyribonucleic acid transformation in *Bacillus subtilis*. *J. Bacteriol.* 86, 785-796

Niu, W., Y. Kim, G. Tau, T. Heyduk, and R.H. Ebright (1996) Transcription activation at class II CAP-dependent promoters: Two interactions between CAP and RNA polymerase. *Cell* 87, 1123-1134

Nuez, B., F. Rojo, and M. Salas (1992) Phage Φ29 regulatory protein p4 stabilizes the binding of RNA polymerase to the late promoter in a process involving direct protein-protein contacts. *Proc. Natl. Acad. Sci. USA* 89: 11401-11405

Ogura, M., H. Yamaguchi, K. Kobayashi, N. Ogasawara, Y. Fujita, and T. Tanaka (2002) Whole-genome analysis of genes regulated by the *Bacillus subtilis* competence transcription factor ComK. *J. Bacteriol.* 184, 2344-2351

O'Neill, M. C., K. Amass and B. de Crombrugghe (1981) Molecular model of the DNA interaction site for the cyclic AMP receptor protein. *Proc. Natl. Acad. Sci U.S.A.* 78, 2213-2217

Parekh, B. S. and G. W. Hatfield (1996) Transcriptional activation by protein-induced bending: evidence for a DNA structural transmission model. *Proc. Natl. Acad. Sci. USA* 93, 1173-1177

Parekh, B. S., S. D. Sheridan, and G. W. Hatfield (1996) Effects of integration host factor and DNA supercoiling on transcription from the *ilv*P_G promoter of *Escherichia coli. J. Biol. Chem.* 271, 20258-20264

Perego, M. and J. A. Hoch (2002) Two-component systems, phosphorelays and regulation of their activities by phosphatases. In *Bacillus subtilis and its closest relatives. From genes to cells.* Sonenshein, A. L., J. A. Hoch and R. Losick (eds). Washington, D. C., American society for microbiology, pp. 473-482

Pérez-Martin, J. and M. Espinosa (1991) The RepA repressor can act as a transcriptional activator by inducing DNA bends. *EMBO J.* 10, 1375-1382

Pérez-Martin, J., F. Rojo and V. de Lorenzo (1994) Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. *Microbiol. Rev.* 58, 268-290

Pérez-Martin, J. and V. de Lorenzo (1997) Clues and consequences of DNA bending in transcription. *Annu. Rev. Microbiol.* 51, 593-628

Persuh, M. K. Turgay, I. Mandic-Mulec and D. Dubnau (1999) The N- and C-terminal domains of MecA recognize different partners in the competence molecular switch. *Mol. Microbiol*. 33, 886-894

Piechowska, M. and M. S. Fox (1971) Fate of transforming deoxyribonuclease in *Bacillus subtilis*. *J. Bacteriol.* 108, 680-689

Piggot, P. J. and J. G. Coote (1976) Genetic aspects of bacterial endospore formation. *Bacteriol. Rev.* 40, 908-962

Priest, F. G. (1993) Systematics and Ecology of *Bacillus*. In *Bacillus subtilis and other Gram-positive bacteria* (Sonenshein, A. L., Hoch, J. A. and Losick, R., eds.), American Society for Microbiology, Washington, D. C., pp. 3-16

Provvedi, R. and D. Dubnau (1999) ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. *Mol. Microbiol.* 31, 271-280

Provvedi, R., I. Chen and D. Dubnau (2001) NucA is required for

DNA-cleavage during transformation of *Bacillus subtilis*. *Mol. Microbiol*. 40, 634-644

Pruss, G.J., and K. Drlica (1989) DNA supercoiling and prokaryotic transcription. *Cell* 56, 521-523

Rao, L., W. Ross, J.A. Appleman, T. Gaal, S. Leirmo, P. J. Schlax, M. T. Record Jr., and R. L. Gourse (1994) Factor independent activation of *rrnB* P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. *J. Mol. Biol.* 235, 1421-1435

Ratnayake-Lecamwasam, M., P. Serror, K. W. Wong and A. L. Sonenshein (2001) *Bacillus subtilis* CodY represses early-stationary phase genes by sensing GTP levels. *Genes and Dev.* 15, 1093-1103

Redfield, R. J., M. R. Schrag and A. M. Dean (1997) The evolution of bacterial transformation: sex with poor relations. *Genetics* 146, 27-38

Roggiani, M. and D. Dubnau (1993) ComA, a phosphorylated response regulator protein of *Bacillus subtilis*, binds to the promoter region of *srfA*. *J. Bacteriol.* 175, 3182-3187

Ross, W., K. K. Gosink, J. Salomon, K. Igarashi, C. Zou, A. Ishihama, K. Severinov, and R. L. Gourse (1993) A third recognition element in bacterial promoters: DNA binding by the α subunit of RNA polymerase. *Science* 262, 1407-1413

Rowe-Magnus, D. A., M. Mencia, F. Rojo, M. Salas, and G. Spiegelman (1998) Transcriptional activation of the *Bacillus subtilis spoIIG* promoter by the response regulator SpoOA is independent of the C-terminal domain of the RNA polymerase alpha subunit. *J. Bacteriol.* 180, 4760-4763

Ryter, A. P., P. Schaeffer and H. Ionesco (1966) Cytologic classification, by their blockage stage, of sporulation mutants of *Bacillus subtilis* Marburg. *Ann. Inst. Pasteur (Paris)* 110, 305-315

Sadler, J. R., H. Sasmor and J. L. Betz (1983) A perfectly symmetric *lac* operator binds the *lac* repressor very tightly. *Proc. Natl. Acad. Sci. U. S. A.* 80, 6785-6789

Sambrook, J., E. F. Fritsch and T. Maniatis (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Serror, P. and A. L. Sonenshein (1996) CodY is required for nutritional repression of *Bacillus subtilis* genetic competence. *J. Bacteriol.* 178, 5910-5915

Singh, R. M. (1972) Number of deoxyribonucleic acid uptake sites in competent cells of *Bacillus subtilis*. *J. Bacteriol.* 110, 266-272

Slack, F. J., P. Serror, E. Joyce and A. L. Sonenshein (1995) A gene

required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon. *Mol. Microbiol.* 15, 689-702

Smith, T. L. and R. T. Sauer (1996) Dual regulation of open-complex formation and promoter clearance by Arc explains a novel repressor to activator switch. *Proc. Natl. Acad. Sci. USA* 93, 8868-8872

Smits, W. K., C. C. Eschevins, K. A. Susanna, S. Bron, L. W. Hamoen and O. P. Kuipers (2005) Stripping *Bacillus*: ComK-autostimulation is responsible for the bistable response in competence development. *Mol. Microbiol.*, in press

Solomon, J. M., R. Magnuson, A. Srivastava and A. D. Grossman (1995) Convergent sensing pathways mediate response to two extracellular competence factors in *Bacillus subtilis*. *Genes Dev.* 9, 547-558

Solomon, J. M. and A. D. Grossman (1996) Who's competent and when: regulation of natural genetic competence in bacteria. *Trends Genet.* 12, 150-155

Solomon, J. M., B. A. Lazazzera and A. D. Grossman (1996) Purification and characterization of an extracellular peptide factor that affects two different developmental pathways in *Bacillus subtilis*. *Genes Dev.* 10, 2014-2024

Stephenson, S., C. Mueller, M. Jiang and M. Perego (2003) Molecular analysis of Phr peptide processing in *Bacillus subtilis*. *J. Bacteriol.* 185, 4861-4871

Straney, D. C. and D. M. Crothers (1985) Intermediates in transcription initiation from the *E. coli lac* UV5 promoter. *Cell* 43, 449-459

Straney, D. C., S. B. Straney, and D. M. Crothers (1989) Synergy between Escherichia coli CAP protein and RNA polymerase in the lac promoter open complex. *J. Mol. Biol*. 206, 41-57

Strauch, M. A. and J. A. Hoch (1993) Transition-state regulators: sentinels of *Bacillus subtilis* post-exponential gene expression. *Mol. Microbiol*. 7, 337-342

Tortosa, P. and D. Dubnau (1999) Competence for transformation: a matter of taste. *Curr. Opin. Microbiol.* 2, 588-592

Towbin, H., T. Staehelin and J. Gordon (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedures and some applications. *Proc. Natl. Acad. Sci. USA*, 76, 4350-4354

Turgay, K., L. W. Hamoen, G. Venema and D. Dubnau (1997) Biochemical characterization of a molecular switch involving the heat shock protein ClpC, which controls the activity of ComK, the competence transcription factor of *Bacillus subtilis*. *Genes Dev.* 11, 119-128

Turgay, K., J. Hahn, J. Burghoorn and D. Dubnau (1998) Competence in *Bacillus subtilis* is controlled by regulated proteolysis of a transcription factor. *EMBO J.* 17, 6730-6738

Van Hijum, S. A. F. T., J. Garcia de la Nava, O. Trelles, J. Kok and O. P. Kuipers (2003) MicroPrep: a cDNA microarray data pre-processing framework. *Applied Bioinformatics* 2, 241-244

Van Hijum, S. A. F. T., R. J. Baerends, H. A. Karsens, A. de Jong, N. E. Kramer, C. J. Albers, J. Kok and O. P. Kuipers (manuscript in preparation) A generally applicable validation scheme for the assessment of factors involved in reproducibility of DNA microarray data.

Van Sinderen, D., S. Withof, H. Boels, and G. Venema (1990) Isolation and characterization of *comL*, a transcription unit involved in competence development in *Bacillus subtilis. Mol. Gen. Genet.* 224, 396-404

Van Sinderen, D. (1994) Regulation of gene expression during competence development in *Bacillus subtilis*. Ph. D. Thesis, University of Groningen, The Netherlands

Van Sinderen, D., A. Ten Berge, B.J. Haijema, L. Hamoen, and G. Venema (1994a) Molecular cloning and sequence of *comK*, a gene required for genetic competence in *Bacillus subtilis*. *Mol. Microbiol.* 11, 695-703

Van Sinderen, D. and G. Venema (1994b) *comK* acts as an autoregulatory control switch in the signal transduction route to compepetence in *Bacillus subtilis*. *J. Bacteriol.* 176, 5762-5770

Van Sinderen, D., A. Luttinger, L. Kong, D. Dubnau, G. Venema, and L. Hamoen (1995) *ComK* encodes the competence transcription factor, the key regulatory protein for competence development in *Bacillus subtilis*. *Mol. Microbiol.* 15: 455-462.

Venema, G., R. H. Pritchard and T. Venema-Schroder (1965) Fate of transforming deoxyribonucleic acid in *Bacillus subtilis*. *J. Bacteriol.* 89, 1250-1255

Wang, H., N. Glansdorff and D. Charlier (1998) The arginine repressor of *Escherichia coli* K-12 makes direct contacts to minor and major groove determinants of the operators. *J. Mol. Biol.* 277, 805-824

Wang, S., R. Cosstick, J. F. Gardner and R. I. Gumport (1995) The specific binding of *Escherichia coli* integration host factor involves both major and minor grooves of DNA. *Biochemistry* 34, 13082-13090

Watson, J. D. and F. H. Crick (1953) The structure of DNA. *Cold Spring Harb. Symp. Quant. Biol.* 18, 123-131

Weinrauch, J., T. Msadek, F. Kunst and D. Dubnau (1991) Sequence and properties of *comQ*, a new competence regulatory gene in *Bacillus subtilis*. *J. Bacteriol.* 173, 5685-5693

Werner, M. H., J. R. Huth, A. M. Gronenborn and G. M. Clore (1995) Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* 81, 705-714

Whipple, F. W., and A. L. Sonenshein (1992) Mechanism of initiation of transcription by *Bacillus subtilis* RNA polymerase at several promoters. *J. Mol. Biol.* 223, 399-414

Workman,C., L. J. Jensen, H. Jarmer, R. Berka, L. Gautier, H. B. Nielsen, H. H. Saxild, C. Nielsen, S. Brunak and S. Knudsen (2002) A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol.* 3, research 0048.1-0048.16

Yuan, G. and S. L. Wong (1995) Regulation of GroE expression in *Bacillus subtilis*: the involvement of the sigma A-like promoter and the roles of the inverted repeat sequence (CIRCE). *J. Bacteriol*. 177, 5427-5433

Zinkel, S. S., and D. M. Crothers (1991) Catabolite activator protein-induced DNA bending in transcription initiation. *J. Mol. Biol.* 219, 201-215

Zomer, A. L. *et al*., manuscript in preparation

Zuber, P., M. M. Nakano and M. A. Marahiel (1993) Peptide antibiotics. In *Bacillus subtilis and other gram-positive bacteria. Biochemistry, physiology and molecular genetics*. Sonenshein, A. L., J. A. Hoch and R. Losick (eds). Washington, D. C., American society for microbiology, pp. 897-916

Appendix 2

List of publications

Susanna, K. A., A. F. van der Werff, C. D. den Hengst, B. Calles, M. Salas, G. Venema, L. W. Hamoen and O. P. Kuipers (2004) Mechanism of transcription activation at the *comG* promoter by the competence transcription factor ComK of *Bacillus subtilis*. *J. Bacteriol.* 186, 1120-1128

Smits, W. K., C. C. Eschevins, K. A. Susanna, S. Bron, L. W. Hamoen and O. P. Kuipers (2005) Stripping *Bacillus*: ComK-autostimulation is responsible for the bistable response in competence development. *Mol. Microbiol.*, in press

Susanna, K. A., F. Fusetti, A.-M. W. H. Thunnissen, L. W. Hamoen and O. P. Kuipers (2005) The C-terminal region of the competence transcription factor ComK of *Bacillus subtilis* is required for transcription activation. Submitted for publication

Susanna, K. A., L. W. Hamoen and O. P. Kuipers (2005) A single, specific thymine mutation in the ComK-binding site severely decreases binding and transcription activation by the competence transcription factor ComK of *Bacillus subtilis*. Submitted for publication

Addendum 1

Nederlandse samenvatting

Achtergrond van het promotie-onderzoek

In het hier beschreven promotie-onderzoek is gewerkt met de staafvormige bodembacterie *Bacillus subtilis*. *B. subtilis* staat bekend om zijn geweldige aanpassingsvermogen aan verslechterende leefomstandigheden. Een mooi voorbeeld hiervan is competentie, een differentiatieproces dat de bacterie in staat stelt om DNA uit zijn omgeving op te nemen en dit vervolgens in te bouwen in zijn eigen genoom. Hierdoor kan de cel nieuwe genetische eigenschappen verkrijgen, die het mogelijk kunnen maken om de verslechterde condities te overleven.
Competentie-ontwikkeling moet op het juiste moment gestart worden. Dit vereist een strenge regulatie. Competentie komt in slechts tien procent van de cellen van een bacterie-populatie voor en is afhankelijk van de aanwezigheid van de activator ComK. Tijdens de exponentiële groeifase wordt de hoeveelheid ComK in de cel gereguleerd. Transcriptie van *comK* wordt geremd door AbrB, CodY en Rok. Aangezien deze controle nooit helemaal sluitend is, wordt er toch een geringe hoeveelheid ComK geproduceerd. Dit wordt gebonden door MecA en afgebroken door het ClpCP-protease-complex. Aan het begin van de stationaire groeifase, stopt de repressie van *comK* door AbrB en CodY, als reactie op de verminderde beschikbaarheid van voedingsstoffen. De toenemende cel-dichtheid in zijn omgeving wordt door de cel opgemerkt via een twee-componenten systeem, resulterend in de productie van ComS. ComS is een klein eiwit dat ComK kan bevrijden uit het protease-complex. Het vrije ComK activeert vervolgens de transcriptie van zijn eigen gen en gaat daarbij de remming door Rok tegen. De ComK-concentratie neemt snel toe en de transcriptie van de andere ComK-afhankelijke genen wordt geactiveerd. Uiteindelijk leidt dit tot de synthese van het DNA-bindings, -opname en -integratiesysteem.
ComK-afhankelijke genen worden gekenmerkt door de aanwezigheid van een ComK-bindingsplaats, een K-box, voor het gen. Deze bestaat uit twee AT-boxen (AAAA-$(N)_5$-TTTT), gescheiden door 8, 18 of 31 baseparen. De AT-boxen bevinden zich aan dezelfde kant van de DNA-helix met, afhankelijk van de lengte van het DNA-fragment, een interval van twee, drie of vier windingen van de helix tussen hun beginposities. ComK bindt als een tetrameer, bestaande uit twee dimeren. De binding van een ComK-tetrameer induceert een buiging in het DNA.

Samenvatting van het promotie-onderzoek

Het hier beschreven onderzoek concentreerde zich vooral op de werking van ComK. Belangrijke vragen waren bijvoorbeeld: Wat is de rol van ComK in het transcriptie-activeringsproces? Zijn er functionele domeinen in het ComK-eiwit te onderscheiden? Welke elementen in een K-box zijn cruciaal voor de activiteit? Kan ComK van *B. subtilis* transcriptie reguleren in een andere bacteriesoort die veel K-boxen bevat, zoals *Lactococcus lactis*? In deze samenvatting worden in het kort de belangrijkste resultaten per hoofdstuk behandeld, waarna ter afsluiting de voornaamste conclusies van dit onderzoek en enkele suggesties voor aanvullend onderzoek beschreven worden.

De rol van ComK in transcriptie-activering

Gentranscriptie wordt uitgevoerd door RNA polymerase (RNAP) en bestaat uit verschillende stappen, beginnend met de binding van RNAP aan de promoter en, via een aantal tussenstappen, eindigend met de voltooiing van het transcript. Vaak wordt transcriptie geactiveerd door een transcriptie-activator, die in principe elk van de genoemde stappen kan stimuleren. In **hoofdstuk 2** is de rol van ComK in het transcriptie-activeringsproces onderzocht, waarbij het ComK-afhankelijke gen *comG* als model gebruikt is. Gelretardaties toonden aan dat RNAP wel kan binden in afwezigheid van ComK, maar dat RNAP-binding sterk gestimuleerd wordt door ComK. Daarnaast bleek ook de aanwezigheid van een extra DNA-fragment, gelegen voor de K-box, noodzakelijk te zijn om RNAP-binding te stabiliseren. Waarschijnlijk maakt de buiging van het DNA, geïnduceerd door ComK-binding, stabiliserende interacties mogelijk tussen het DNA gelegen voor de K-box en RNAP.
De K-box van het model-gen *comG* wijkt af van de standaard K-box door de aanwezigheid van een extra AT-box, gelegen voor de normale K-box. Voor het aantonen van transcriptie *in vitro* (in reacties buiten de cel), is de aanwezigheid van deze derde AT-box noodzakelijk. In de cel is de extra AT-box echter niet nodig, aangezien ComK daar transcriptie kan activeren van zowel *comG* voorafgegaan door een standaard K-box als ook van de andere ComK-afhankelijke genen, die meestal een K-box hebben die bestaat uit twee AT-boxen.

RNAP-binding is de eerste stap in transcriptie. In **hoofdstuk 2** is ook de rol van ComK in de andere stappen onderzocht, maar deze bleken ComK-onafhankelijk te zijn. De voornaamste rol van ComK in transcriptie-activering is dus het stabiliseren van RNAP-binding door interacties met het DNA voor de K-box mogelijk te maken. Ondanks afwijkingen van de K-box van *comG* ten opzichte van de standaard, mag aangenomen worden dat ComK transcriptie van andere afhankelijke genen op vergelijkbare manier stimuleert.

Functionele domeinen in het ComK-eiwit

Een kenmerk van ComK is dat het eiwit betrokken is bij verschillende interacties tijdens competentie-regulatie: i) eiwit-eiwit interacties, zowel met andere eiwitten (MecA) als tussen ComK-eiwitten onderling voor oligomerisatie en ii) eiwit-DNA interacties bij binding aan een K-box. De betrokkenheid van ComK bij diverse interacties suggereert de aanwezigheid van specifieke functionele domeinen in het eiwit. In **hoofdstuk 3** is gebruik gemaakt van een aantal ComK-mutanten om verschillende functionele domeinen te karakteriseren.
De best gekarakteriseerde mutant, ComKΔC25 mist 25 aminozuren van het C-terminale uiteinde en bleek geen transcriptie meer te kunnen activeren, maar nog wel DNA te kunnen binden. DNA-binding gebeurde wel met dezelfde affiniteit als het oorspronkelijke ComK, maar als dimeer in plaats van als tetrameer. Dit duidt erop dat het tetramerisatiedomein van ComK zich aan het C-terminale einde bevindt. Een verlies van tetramerisatie, leidt tot een afname van DNA-buiging, die door binding van een ComK-tetrameer geïnduceerd wordt. Zoals beschreven in **hoofdstuk 2** is deze buiging cruciaal voor het stabiliseren van RNAP-binding en dus voor het activeren van transcriptie.
Een andere interessante eiwitmutant is ComKΔN9, waarbij negen aminozuren vanaf de N-terminus verwijderd zijn. Deze variant is relatief actiever dan het oorspronkelijke ComK in transcriptie-activering. In *B. subtilis* is de expressie van ComKΔN9 echter slechts 15-20% van het niveau van het originele ComK. Dit suggereert dat de N-terminale regio van ComK betrokken is bij de regulatie van de productie en/of stabiliteit van het eiwit. De lengte van deze regio zou een compromis kunnen zijn tussen de regulatie van ComK-expressie en een maximale transcriptie-activering.

Cruciale elementen voor activiteit van een K-box

De consensus sequentie van een K-box kent een grote natuurlijke variatie. Een computeranalyse van het genoom van *B. subtilis* laat de aanwezigheid zien van meer dan duizend potentiële K-boxen. Circa 30% van deze K-boxen ligt in een promoter-gebied, maar slechts 8% is actief in regulatie door ComK. Kennelijk zijn dus niet alle variaties toegestaan. **Hoofdstuk 4** van dit proefschrift beschrijft een studie waarin K-boxen gemuteerd zijn om cruciale elementen voor de activiteit van de box aan te tonen. De opvallendste natuurlijke variatie is de lengte van het interval tussen de AT-boxen: 8, 18 of 31 baseparen voor respectievelijk een klasse I, II of III K-box. Het eerste onderzoek in **hoofdstuk 4** toonde aan dat klasse III K-boxen minder actief zijn dan klasse I en II boxen. Naast de lengte van het interval, kan ook het A/T-percentage van het DNA in het interval van belang zijn, aangezien A/T-baseparen flexibeler zijn dan G/C's. Een hoog A/T-percentage zou noodzakelijk kunnen zijn voor het induceren van een buiging van het DNA door ComK. Verlaging van het A/T-percentage van 60% naar 40%, bleek echter weinig effect te hebben op de activiteit van de K-box. Het tweede onderzoek van **hoofdstuk 4** bestond uit het systematisch introduceren van mutaties in de consensus sequentie van een AT-box. Om te beginnen is de K-box voor het *comG*-gen, geïdealiseerd tot een perfecte consensus sequentie. Op elke positie van de eerste AT-box is vervolgens een G geïntroduceerd in plaats van een A of een T. Het opvallendste resultaat was dat de mutatie van de tweede T in de AT-box leidde tot een afname van ComK-binding en transcriptie tot slechts 30% van het niveau van de geïdealiseerde K-box. Wanneer ook de T_2-positie in de andere AT-box gemuteerd werd, was er nauwelijks nog activiteit meetbaar. Een vergelijking van de sequenties van alle bekende, actieve K-boxen, bevestigde het belang van de T_2-posities. Geen van deze 88 K-boxen bleek een dubbele T_2-mutatie te bevatten. Kennelijk is dus tenminste één ongemuteerde T_2-positie noodzakelijk voor activiteit van een K-box. Een mogelijke verklaring zou een sterk coöperatieve binding van de twee ComK-dimeren kunnen zijn. De eerste ComK-dimeer zou kunnen binden aan de ongemuteerde AT-box, waarna binding van de tweede dimeer kan volgen, zelfs aan een T_2-gemuteerde AT-box. Wanneer beide AT-boxen een T_2-mutatie bevatten, is een stabiele dimeer-binding onmogelijk, waardoor de K-box inactief is.

Introductie van *B. subtilis* ComK in een andere bacterie

K-boxen komen veel voor, zeker in bacteriën met een AT-rijk genoom. Ook *Lactococcus lactis* bevat veel K-boxen. Voor het onderzoek naar functionele domeinen in ComK (**hoofdstuk 3**), is *B. subtilis comK* tot expressie gebracht in *L. lactis* met behoud van activiteit, zoals aangetoond door transcriptie vanaf de *B. subtilis comG*-promoter. Deze observatie roept de vraag op of ComK ook transcriptie van *L. lactis*-genen kan reguleren door te binden aan de aanwezige K-boxen in het genoom. DNA-microarray studies toonden aan dat *comK*-expressie de transcriptie van circa 200 genen beïnvloedt, voor het merendeel remmend. De meeste effecten waren indirect: slechts 12% van de geactiveerde en 27% van de geremde genen werd voorafgegaan door een K-box. Het relatief grote aantal directe remmende effecten in *L. lactis* is opvallend en is tegengesteld aan de situatie in *B. subtilis*, waar ComK eigenlijk uitsluitend als activator werkt. Het zou kunnen dat ComK bindt aan de beschikbare K-boxen en daardoor de normale transcriptie in *L. lactis* verstoort. Deze hypothese werd bevestigd door een DNA-microarray studie met de ComK-variant ComKΔC25, die dezelfde genen beïnvloedde als oorspronkelijk ComK. In dit geval kunnen directe effecten alleen veroorzaakt worden door binding van ComK, aangezien de mutant geen transcriptie kan activeren.
Een zeer interessante conclusie uit deze studie is dat de functie van een eiwit in een gastbacterie niet automatisch hetzelfde is als in de oorspronkelijke bacterie. De werking van een regulator is in belangrijke mate afhankelijk van de condities en factoren, zoals RNAP, in de cel waarin het eiwit tot expressie gebracht wordt.

Suggesties voor verder onderzoek

Het hier beschreven promotie-onderzoek richtte zich vooral op een functionele analyse van de domeinen en de werking van het ComK-eiwit. De rol van ComK in het transcriptie-activeringsproces is aangetoond in **hoofdstuk 2**, waarmee dit werk afgesloten is. Het onderzoek naar de positie van functionele domeinen in ComK en cruciale elementen in een K-box laat meer vragen open en is dan ook nog niet afgerond. Hoewel **hoofdstuk 3** en **4** de locatie van het tetramerisatiedomein van ComK en het belang van de T_2-posities in een K-box voor stabiele ComK-binding beschrijven, is

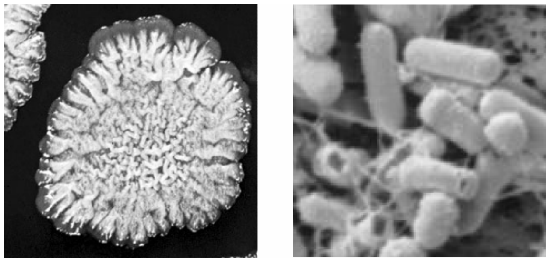
het mechanisme van tetramerisatie nog onduidelijk, evenals de exacte rol van de T_2's. Een eiwitkristallisatie-studie van ComK, bij voorkeur gebonden aan DNA, zou meer duidelijkheid kunnen verschaffen. Indien succesvol, kan een dergelijke studie helpen om tot een compleet beeld te komen van de verschillende functionele domeinen in het eiwit en hun interacties. Naast de reeds bekende locatie van het tetramerisatiedomein, zouden dan ook het dimerisatie- en het DNA-bindingsdomein aangetoond kunnen worden. Verder kan een dergelijk onderzoek helpen om individuele aminozuren en baseparen aan te wijzen, die betrokken zijn bij verschillende interacties. In samenwerking met de groep van biofysische chemie aan de Rijksuniversiteit Groningen is een begin gemaakt met eiwitkristallisaties, maar tot dusver zijn deze niet succesvol gebleken. Gezien de gedetailleerde informatie over het eiwit en zijn interacties, die verkregen kan worden uit een driedimensionaal beeld, is het aan te raden om in de toekomst de aandacht toch vooral te richten op het verkrijgen van de kristalstructuur van ComK-DNA complexen.

Addendum 2

Voor de leek

Bacteriën, organismen met een enorme verscheidenheid

Bacteriën worden algemeen beschouwd als de eenvoudigste vorm van leven op aarde. Het zijn kleine (**micro-**)**organismen**, die vaak niet groter zijn dan een paar duizendste millimeter (figuur 1). Bacteriën komen overal ter wereld voor en er bestaat een grote verscheidenheid aan soorten. De meeste bacteriën leven onder tamelijk gematigde condities, bijvoorbeeld in de grond, in het water of op planten. Sommige soorten zijn echter aangepast aan een verblijf onder zeer moeilijke omstandigheden, waar verder nauwelijks leven mogelijk is. Ze komen voor rondom vulkanen, in zwavelbronnen, op sneeuwvelden, onder extreem hoge druk of in extreem zout water. Vanwege hun extreme leefomgeving, worden deze bacteriën ook wel **extremofielen** genoemd.



Figuur 1. Met het blote oog zijn bacteriën te zien als een kolonie bestaande uit miljoenen cellen (links). Met een sterk vergrotende microscoop zijn de individuele bacteriecellen waarneembaar (rechts)

Dichter bij huis treffen we ook veel bacteriën aan, bijvoorbeeld in ons voedsel of in ons eigen lichaam, dat meer micro-organismen huisvest dan er mensen op aarde zijn. De meeste soorten zijn ongevaarlijk en vaak zeer nuttig. Ze worden ook wel **commensalen** (letterlijk: tafelgenoten) genoemd. Deze micro-organismen leven op ons huidoppervlak en op alle slijmvliezen die toegankelijk zijn vanuit de buitenwereld, zoals de mond, de neus en het maag-darmkanaal, waar ze belangrijk zijn voor een goed verloop van de spijsvertering. Een bijkomend voordeel van de aanwezigheid van onschadelijke bacteriën is dat ze het lichaam beschermen tegen schadelijke soorten (**pathogenen**), door de leefomgeving bezet te houden. Soms lukt het pathogene bacteriën echter wel om het lichaam binnen te dringen. Het afweersysteem treedt dan in werking en zal ze meestal snel onschadelijk kunnen maken. Wanneer

dit niet lukt, kan er een infectie optreden. Enkele bekende voorbeelden van ziektes met een bacteriële oorzaak zijn kinkhoest, de veteranenziekte en tuberculose. Ook in ons voedsel komen zowel schadelijke als nuttige bacteriën voor. Zeer berucht zijn de *Listeria* bacteriën, die soms worden aangetroffen in producten die bereid zijn met rauwe melk en de *Salmonella* soorten die vaak op eieren en in kipfilet voorkomen. Daarnaast zijn er ook veel bacteriesoorten die juist gebruikt worden om voedsel te maken, zoals de melkzuurbacteriën die van groot belang zijn voor de productie van verschillende zuivelsoorten als yoghurt, kaas en karnemelk.
De naam van een bacterie wordt altijd cursief geschreven, bijvoorbeeld *Bacillus subtilis* of *Lactococcus lactis*. *Bacillus* en *Lactococcus* geven aan tot welke groep een bacterie behoort, terwijl *subtilis* en *lactis* de specifieke soort aanduiden. Bij herhaling van de naam, wordt het eerste gedeelte alleen nog met de beginletter aangeduid, dus *B. subtilis* en *L. lactis*.

Indeling van levensvormen

De enorme diversiteit aan bacteriesoorten blijkt wel uit de indeling van alle levensvormen op aarde in verschillende domeinen. Van oudsher werd al het leven geclassificeerd in vijf rijken: planten, dieren, schimmels, protozoa (kleine eencellige organismen) en bacteriën. Moderne technieken hebben echter aangetoond dat planten, dieren, schimmels en protozoa relatief zo sterk verwant zijn dat ze allemaal samengevoegd kunnen worden tot één rijk, de **eukaryoten**. De grote groep bacteriën moet daarentegen juist gesplitst worden in twee rijken: de **archaea** en de **bacteriën**, die evenveel van elkaar verschillen als ieder van hen afzonderlijk verschilt van de eukaryoten.

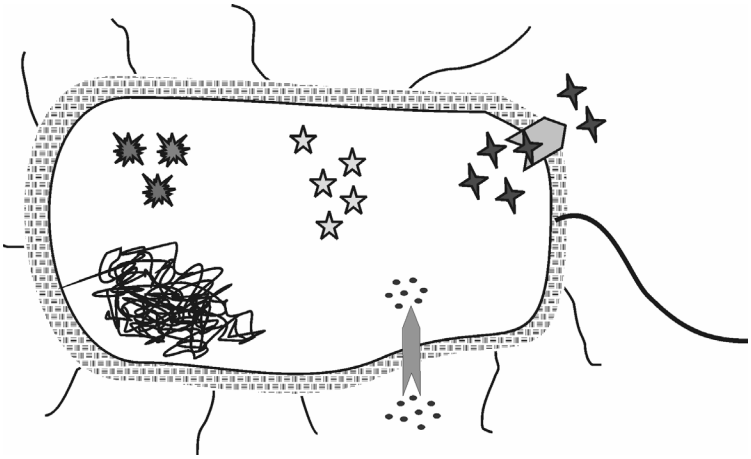
Bacteriën in wetenschappelijk onderzoek

Bacteriën zijn interessante levensvormen voor onderzoekers. De belangrijkste reden hiervoor is dat ze **eencellig** zijn en dus relatief eenvoudig. Ter vergelijking: schattingen van het aantal cellen in een menselijk lichaam lopen uiteen van tien tot honderd biljoen. Een ander groot voordeel van bacteriën is hun korte generatietijd. Onder gunstige omstandigheden kan een bacteriële cel

zich in twintig minuten delen. Veranderingen in de erfelijke eigenschappen kunnen dan ook snel doorgegeven worden. Hierdoor is het mogelijk om de effecten van aanpassingen van het erfelijk materiaal van een cel binnen korte tijd te bepalen. Zelfs de meest eenvoudige meercellige levensvormen hebben een generatietijd die vele malen langer is dan die van bacteriën, om maar niet te spreken van de gemiddeld dertig jaar die het duurt voordat er een nieuwe generatie mensen is ontstaan.

De cel, de basis van het leven

De basis van alle levensvormen wordt gevormd door een cel. Maar wat is een cel eigenlijk? Een **cel** kan gedefinieerd worden als een afgesloten compartiment dat het erfelijk (**genetisch**) materiaal (**DNA**) van het organisme bevat en de machinerie om afschriften van het DNA te maken en deze te vertalen in de uitvoerders (**eiwitten**), die de verschillende processen in een cel voor hun rekening nemen. Een schematische weergave van een bacteriële cel is afgebeeld in figuur 2.



Figuur 2. Schematische tekening van een bacteriële cel
De cel is omgeven door een afscheiding, in dit geval bestaande uit een membraan met daaromheen de celwand. In de cel bevinden zich het DNA, als een kluwen, en de verschillende eiwitten, weergegeven door sterretjes. Transport tussen de cel en zijn omgeving kan plaatsvinden via openingen in de membraan en de celwand, aangegeven met pijltjes. De haarvormige uitsteeksels aan de buitenkant zorgen voor mobiliteit van de cel.

Wanneer een cel zich deelt, wordt het DNA gekopieerd en verdeeld over de moedercel en de nieuwgevormde dochtercel, zodat er twee identieke cellen ontstaan met dezelfde erfelijke eigenschappen. Elke cel wordt omsloten door een afscheiding, bestaande uit een of meerdere **membranen** en soms ook een **celwand**. Op deze manier wordt de cel beschermd tegen invloeden van buitenaf. Het is echter geen absolute barrière, aangezien de cel wel moet leven in zijn omgeving en daar ook op moet kunnen reageren. Om signalen met zijn omgeving uit te wisselen, maakt de cel gebruik van openingen in de membraan en de celwand waardoor transport van verschillende componenten vanuit de cel naar buiten en vanuit de omgeving naar de cel kan plaatsvinden. De inhoud van een cel wordt het **cytoplasma** genoemd. Dit bestaat uit een waterige oplossing waarin zich de eiwitten bevinden. Eukaryote en bacteriële cellen verschillen qua organisatie in de cel, vooral wat de positie van het DNA betreft. In bacteriën bevindt het DNA zich los in het cytoplasma, terwijl eukaryote cellen een apart compartiment, de celkern, hebben om het DNA in op te slaan.

DNA, de informatiebron van de cel

Het DNA bestaat uit twee strengen die om elkaar heen gewonden zijn en samen een **dubbele helix** vormen. Elke streng is opgebouwd uit bouwstenen (**basen**). Er zijn vier verschillende basen: adenine (**A**), thymine (**T**), cytosine (**C**) en guanine (**G**), die onderling **baseparen** kunnen vormen. Tegenover een A op de ene DNA-streng zit altijd een T op de andere streng en tegenover een C zit altijd een G. Hierdoor zijn de twee DNA-strengen van een helix complementair aan elkaar. Een schematische weergave van de complementaire baseparing in een DNA-helix is weergegeven in figuur 3.

Figuur 3. Overzicht van de structuur van DNA
De twee DNA-strengen vormen een dubbele helix. De verticale streepjes geven de baseparen aan die beide strengen verbinden. De letters tonen een voorbeeld van complementaire baseparing.

De volgorde van de verschillende basen in een DNA-streng (de **sequentie**), bepaalt de eigenschappen die in het DNA opgeslagen liggen. Het begin en het einde van de DNA-helix zijn bij een bacterie aan elkaar verbonden, zodat het DNA uit één cirkelvormig **chromosoom** bestaat. Ter vergelijking: de mens heeft 46 lineaire chromosomen in elke cel. Het totale DNA in een cel wordt het **genoom** genoemd en van veel organismen is inmiddels de gehele sequentie van het genoom bepaald.

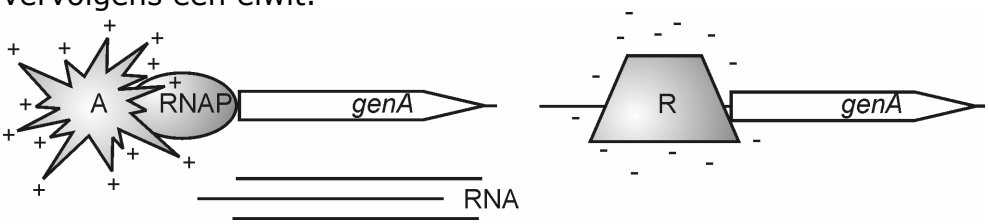
Van DNA naar eiwit

Op een chromosoom zijn verschillende onderdelen (**genen**) te onderscheiden, die coderen voor een bepaalde eigenschap van de cel. Een gen is te herkennen aan een duidelijk start- en stopsignaal. Op basis van de informatie in een gen wordt een specifiek eiwit gemaakt. Meestal zijn de naam van het gen en van het bijbehorende eiwit hetzelfde. Om ze te onderscheiden wordt de gennaam cursief geschreven (*genA*) en de naam van het eiwit met een hoofdletter (GenA). Om een eiwit te kunnen maken, wordt eerst een afschrift van het DNA van een gen gemaakt (**transcriptie**). Dit afschrift, het **RNA**, lijkt qua structuur op een DNA-streng en bevat dezelfde informatie, maar wordt slechts gemaakt als een tijdelijke kopie. Wanneer er voldoende eiwit is, worden er geen nieuwe afschriften meer gemaakt en wordt het nog aanwezige RNA weer afgebroken. Net als DNA en RNA bestaan eiwitten uit losse bouwstenen (**aminozuren**). Er bestaan twintig verschillende aminozuren en de volgorde en aanwezigheid van de diverse aminozuren bepaalt de eigenschappen van een eiwit. Op basis van de informatie die opgeslagen ligt in het RNA, worden de juiste aminozuren door de eiwitsynthese-machinerie aan elkaar gekoppeld en wordt het eiwit gevormd (**translatie**).

Hoe reguleert een cel de productie van eiwitten?

Eiwitten zijn betrokken bij de meeste processen die in een cel plaatsvinden en kunnen verschillende functies hebben, variërend van het reguleren van processen tot het daadwerkelijke uitvoeren ervan. Sommige eiwitten zijn altijd in de cel aanwezig, terwijl andere eiwitten slechts tijdelijk nodig zijn. Of en wanneer een eiwit

nodig is, hangt af van de omstandigheden waaronder de cel leeft. Zijn er nog voldoende voedingsstoffen? Gaat de cel zich delen? Moet er een bepaald proces opgestart worden als gevolg van veranderde condities? Een cel ontvangt continu signalen over zijn leefomgeving en reageert hierop door de productie van eiwitten, de **eiwitexpressie**, te reguleren. Hiervoor is controle nodig van de transcriptie van het gen dat codeert voor het gevraagde eiwit (**transcriptionele regulatie** of **genregulatie**). Om een afschrift van een gen te maken is het eiwit **RNA-polymerase** nodig. RNA-polymerase (**RNAP**) bindt aan de **promoter**, een vaste sequentie in het DNA voor de start van het gen. Vervolgens beweegt RNAP zich langs het gen en bouwt, in de juiste volgorde, basen in de RNA-kopie in. Sommige genen staan altijd aan, maar in de meeste gevallen wordt transcriptie gereguleerd door **repressor**- en **activator**-eiwitten (figuur 4). Een repressor bindt meestal op de plaats van het RNA-polymerase en verhindert zo RNAP-binding. Er kan dan geen RNA-kopie gemaakt worden en het gen staat uit. Een activator bindt vaak naast RNAP en helpt bij binding en/of activering van RNAP. Er wordt dan wel een afschrift gemaakt en vervolgens een eiwit.



Figuur 4. Overzicht van de werking van een activator en een repressor
Een activator (A) bindt meestal naast het RNAP en stimuleert binding en/of activering van RNAP. Er worden dan RNA-kopieën gemaakt. Een repressor (R) bindt meestal op de plaats van RNAP en verhindert zo RNAP- binding. Er worden dan geen afschriften gemaakt en het gen staat uit.

Hoe herkent een regulator de juiste genen?

Er zou geen sprake zijn van specifieke genregulatie wanneer een activator of repressor willekeurig aan het DNA zou binden en de naastgelegen genen zou reguleren. Hoe herkent een eiwit waar hij moet binden en welke genen hij moet reguleren? Genregulatie is afhankelijk van de aanwezigheid van een **bindingsplaats** voor de regulator. Bindingsplaatsen worden gekenmerkt door een be-

paalde volgorde in het DNA en zijn voor alle eiwitten anders. Elk eiwit heeft zijn eigen **herkenningssequentie**. Eiwit A herkent bijvoorbeeld de volgorde GGCAATT in het DNA, bindt hieraan en reguleert de transcriptie van het naastgelegen gen, terwijl eiwit B een andere herkenningssequentie heeft, bijvoorbeeld TTCTAAAG. Eiwit A herkent de bindingsplaats voor eiwit B niet en vice versa. Ze reguleren alleen de transcriptie van genen die voorafgegaan worden door hun eigen, specifieke bindingsplaats. Naast verschillen in herkenningssequenties zijn er ook verschillen in binding van eiwitten. Sommige binden als een enkel eiwit (een **monomeer**) aan het DNA, terwijl andere eiwitten als een complex van meerdere eiwit-kopieën binden. De algemene benaming voor een complex gevormd door meerdere monomeren is een **oligomeer**.

Regulatie van de activiteit van een eiwit

Naast de controle op de productie van eiwitten door middel van transcriptionele regulatie, kan ook de activiteit van een reeds geproduceerd eiwit gereguleerd worden. Deze vorm van regulatie heet **post-translationele regulatie**. Een mogelijkheid is bijvoorbeeld het binden van een doel-eiwit door een regulerend eiwit, waardoor het eerste eiwit afgebroken of in een inactieve vorm gehouden wordt. Door een combinatie van transcriptionele en post-translationele regulatie wordt de productie en activiteit van eiwitten in een cel gecontroleerd en ontstaat er een complex netwerk dat regulatie en uitvoering van verschillende processen in de cel met elkaar verbindt. Dit resulteert in een functionele cel die kan groeien, delen en contact kan onderhouden met zijn omgeving.

Bacillus subtilis, een modelbacterie

In dit promotie-onderzoek is gewerkt met de bacterie ***Bacillus subtilis***, een staafvormige bodembacterie, die algemeen beschouwd wordt als het model-organisme voor een grote groep bacteriën. *B. subtilis* cellen zijn ongeveer 2 micrometer lang (een micrometer is 0,001 millimeter) en zijn te zien in figuur 1. *B. subtilis* is een prettige bacterie om mee te werken, aangezien hij volkomen onschadelijk is. Daarnaast is de bacterie eenvoudig te kweken in het laboratorium en is het relatief gemakkelijk om ver-

anderingen in het DNA aan te brengen. De genoomsequentie van *B. subtilis* is bekend en bevat ongeveer vier miljoen baseparen. Circa 60% van alle basen bestaat uit een A of een T. *B. subtilis* wordt daarom gerekend tot de A/T-rijke bacteriën. Het genoom kan onderverdeeld worden in ruim 4000 genen. Van slechts een gedeelte van deze genen is de functie bekend.
De familie waartoe *B. subtilis* behoort, bevat veel bacteriesoorten. De meeste zijn totaal onschadelijk en soms zelfs zeer nuttig. In de wasmiddelen-industrie worden *Bacillus*-soorten bijvoorbeeld op uitgebreide schaal ingezet voor het produceren en uitscheiden van grote hoeveelheden eiwitten. Sommige *Bacillus* soorten worden zelfs gegeten. Een traditioneel Japans gerecht is natto, een sojaproduct dat bereid wordt met behulp van *Bacillus*-bacteriën. De *Bacillus* familie bevat echter ook een paar pathogene bacteriën, waarvan *Bacillus anthracis*, de veroorzaker van het dodelijke miltvuur, de bekendste is.

B. subtilis **bezit een geweldig aanpassingsvermogen**

B. subtilis is vooral bekend vanwege zijn fantastische vermogen om zich aan te passen aan wisselende groeicondities. Onder goede omstandigheden deelt de bacterie zich elke twintig à dertig minuten, resulterend in een **exponentiële groei** van het aantal cellen. Wanneer de leefomstandigheden verslechteren, bijvoorbeeld door de afname van beschikbare voedingsstoffen, vlakt de groei van de populatie af. Deze fase wordt de **stationaire groeifase** genoemd. Aan het begin van deze fase, start *B. subtilis* een aantal **aanpassingsprocessen** op, die de cel in staat kunnen stellen om concurrenten uit te schakelen of een grotere verscheidenheid aan voedingsbronnen te benutten. Als de slechte omstandigheden aanhouden, neemt de cel zijn toevlucht tot twee meer drastische aanpassingsprocessen: **competentie** en uiteindelijk **sporulatie**. Een competente cel kan DNA opnemen uit zijn omgeving en dit inbouwen in zijn eigen genoom. Hierdoor kan de bacterie nieuwe eigenschappen verkrijgen, die het mogelijk kunnen maken om de verslechterde condities te overleven. Sporulatie resulteert in de vorming van een spore, een slapende vorm van de cel, die gedurende lange tijd en onder zeer moeilijke condities kan overleven. Zodra de omstandigheden weer gunstig zijn, kan de spore ontkiemen en een nieuwe *B. subtilis*-populatie starten.

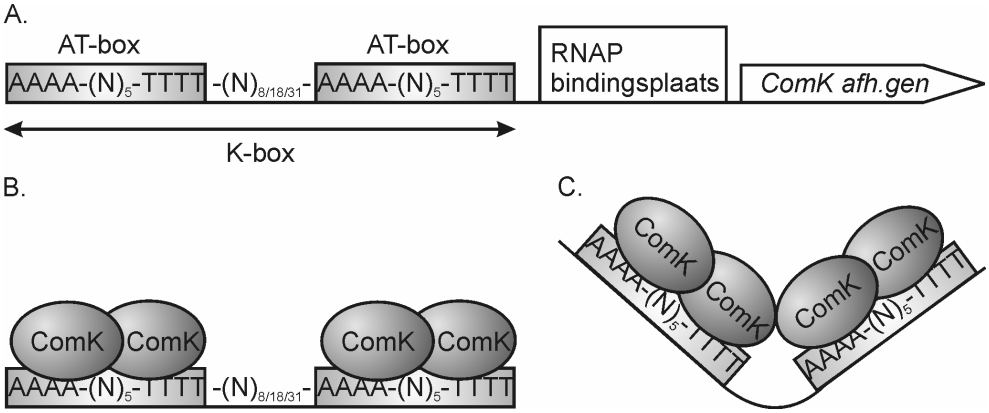
Regulatie van competentie ontwikkeling

Om ervoor te zorgen dat de aanpassingsprocessen pas opgestart worden wanneer de omstandigheden verslechteren, is strenge en vaak complexe regulatie nodig. Een mooi voorbeeld hiervan is de regulatie van competentie. Competentie komt slechts voor in tien procent van de cellen en is afhankelijk van de aanwezigheid van de transcriptie-activator **ComK**. ComK stimuleert transcriptie van ComK-afhankelijke genen. Dit resulteert in de productie van eiwitten die betrokken zijn bij de opname van DNA uit de omgeving en het inbouwen van dit DNA in het genoom van de cel. Competentie mag echter pas opgestart worden in de stationaire groeifase en zeker niet eerder, aangezien het dan grote schade aan de cellen toe kan brengen. Het is dus van groot belang om de productie en de activiteit van ComK tijdens de exponentiële groeifase te controleren. Dit gebeurt door zowel transcriptionele als post-translationele regulatie. De transcriptionele regulatie bestaat uit de controle van transcriptie van het *comK*-gen door een aantal repressor-eiwitten. Toch wordt er dan nog wel een beetje ComK-eiwit gemaakt, omdat de controle niet helemaal sluitend is. Het geproduceerde ComK wordt onschadelijk gemaakt door post-translationele regulatie, waarbij ComK gebonden wordt door een ander eiwit, MecA, en vervolgens wordt afgebroken. Wanneer de cel signalen krijgt dat de omstandigheden verslechteren, wordt de controle van ComK opgeheven. ComK is dan vrij in de cel en kan transcriptie activeren. Om snel een grote hoeveelheid ComK te krijgen, activeert ComK ook de transcriptie van zijn eigen gen.

De herkenningssequentie en binding van ComK

ComK stimuleert competentie-ontwikkeling door te binden aan het DNA voor de start van ComK-afhankelijke genen. ComK bindt aan een ComK-bindingsplaats, ook wel een **K-box** genoemd en herkent deze box aan zijn herkenningssequentie (figuur 5). In het geval van ComK is er sprake van een complexe herkenningssequentie, namelijk twee keer AAAA-$(N)_5$-TTTT (een **AT-box**), gescheiden door een DNA fragment met een willekeurige sequentie. Er zijn drie verschillende lengtes van dit fragment mogelijk, 8, 18 of 31 baseparen. Een N in de herkenningssequentie geeft aan dat hier elk van de vier voorkomende baseparen toegestaan is. In

totaal binden vier ComK-eiwitten (een **tetrameer**) aan een K-box. De tetrameer kan onderverdeeld worden in twee groepjes van twee ComK-eiwitten (een **dimeer**). Een belangrijk kenmerk van ComK-binding is dat het vergezeld gaat van een buiging van het DNA, die van belang is voor de functionaliteit van ComK.

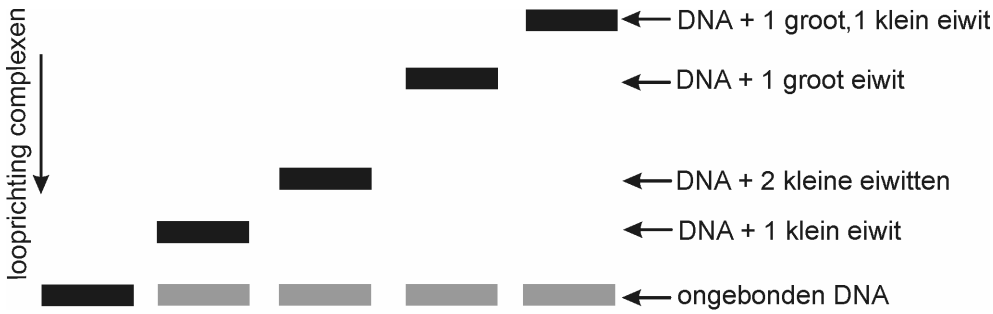


Figuur 5. ComK-binding aan een K-box
(A). Een K-box bestaat uit twee AT-boxen en ligt voor de RNAP bindingsplaats en het ComK-afhankelijke gen. (B). ComK bindt als twee dimeren, die een tetrameer vormen. Daarbij wordt het DNA gebogen (C).

Hoe activeert ComK transcriptie?

Bovenstaande informatie was reeds bekend vanuit onderzoek van anderen en vormt de achtergrond waartegen dit proefschrift tot stand gekomen is. Het hier beschreven promotie-onderzoek richtte zich vooral op de functionaliteit en de karakteristieken van het ComK-eiwit en op de cruciale elementen in een K-box. In het volgende gedeelte van dit hoofdstuk zal een samenvatting van de belangrijkste onderzoeksresultaten gegeven worden.
De eerste uitdaging bestond uit het ophelderen van de rol van ComK in het transcriptie-activeringsproces. Transcriptie-activering bestaat uit verschillende stappen die uitgevoerd worden door RNA polymerase, te beginnen bij binding aan de promoter en, via een paar tussenstappen, eindigend met het voltooien van het RNA-afschrift. Een activator kan in elke stap helpen om deze makkelijker te laten verlopen. Om de rol van ComK in transcriptie-activering te onderzoeken, is het ComK-afhankelijke gen *comG* als mo-

del gebruikt. Als eerste werd het effect van ComK op binding van RNAP onderzocht. Om binding van eiwitten aan DNA aan te tonen, wordt gebruik gemaakt van **gel-retardaties**. Bij deze techniek wordt de DNA-regio rondom de start van een gen, in dit geval *comG*, gebruikt. Aan dit DNA worden verschillende eiwitten, in dit geval RNAP en/of ComK toegevoegd, zodat ze kunnen binden. Het mengsel wordt op een gel gebracht, die bestaat uit een netwerk waar DNA en eiwitten doorheen kunnen lopen. De loopsnelheid hangt af van de grootte van het complex. Een klein complex (bijvoorbeeld alleen DNA) zal snel door de mazen lopen, terwijl een groter complex (bijvoorbeeld eiwit-gebonden DNA) langzamer loopt. Op deze manier kunnen verschillende complexen gescheiden worden. Om ze zichtbaar te maken, is een radioactief signaal aan het DNA gekoppeld. Het resultaat is een plaatje zoals weergegeven is in figuur 6.



Figuur 6. Schematische weergave van een gel-retardatie
De tekst (rechts) geeft een mogelijke verklaring voor elk complex.

De onderste banden zijn snel gelopen en bevatten alleen DNA, terwijl de bovenste banden langzamer gelopen zijn en naast DNA ook gebonden eiwit bevatten. Op deze manier is voor de *comG*-promoter aangetoond dat RNAP-binding sterk gestimuleerd wordt door de aanwezigheid van ComK. Deze stimulatie bleek afhankelijk te zijn van de aanwezigheid van een lang stuk DNA, gelegen voor de ComK-bindingsplaats. Zoals eerder gezegd, leidt ComK-binding aan een K-box tot buiging van het DNA. Dit kan contacten mogelijk maken tussen het verderop gelegen DNA en RNA-polymerase, waardoor RNAP-binding gestabiliseerd wordt. De belangrijkste rol van ComK in transcriptie-activering is dus het stabiliseren van RNAP-binding door contacten met het verderop gelegen DNA mogelijk te maken. Het gebonden RNAP kan daarna transcriptie van het *comG*-gen verzorgen. Hoewel het mechanisme al-

leen voor *comG* onderzocht is, is het aannemelijk dat het voor de andere ComK-afhankelijke genen op dezelfde manier werkt en dat ComK transcriptie activeert door RNAP-binding te stimuleleren.

Functionele domeinen in het ComK eiwit

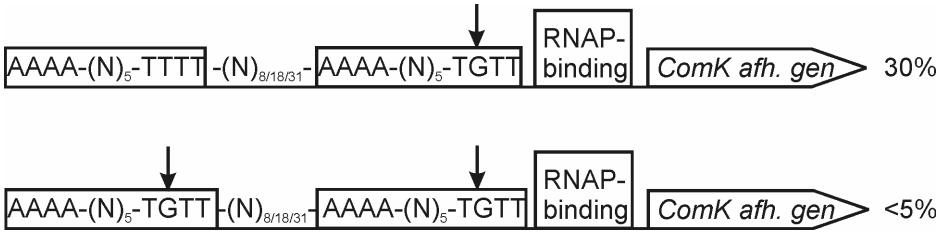
Een van de meest interessante aspecten van de werking van ComK in *B. subtilis* is dat ComK betrokken is bij verschillende **interacties**, zowel met andere eiwitten, als tussen ComK-eiwitten onderling als ook met het DNA van de ComK-bindingsplaats. Vaak kunnen in een eiwit verschillende **domeinen** onderscheiden worden die verantwoordelijk zijn voor het aangaan van een bepaalde interactie. ComK zou bijvoorbeeld moeten beschikken over een **DNA-bindingsdomein**, een **dimerisatie**- en een **tetramerisatiedomein** (voor de vorming van een complex van respectievelijk twee en vier ComK-eiwitten). Deze drie interacties vinden tegelijk plaats en kunnen dus niet door hetzelfde gedeelte van het eiwit verzorgd worden. ComK is een relatief klein eiwit dat bestaat uit 192 aminozuren. Ter vergelijking: RNAP bestaat uit een complex van verschillende eiwitten, waarvan de grootste meer dan 1000 aminozuren bevat. Om verschillende domeinen in ComK aan te tonen, zijn **eiwitmutanten** gemaakt, die een aantal aminozuren missen. De functionaliteit van deze mutanten is getest met verschillende experimenten: gel-retardaties om binding van ComK aan het DNA te laten zien en **betagalactosidase-experimenten** om transcriptie-activering aan te tonen. Bij deze experimenten is een indicator-gen, *lacZ*, onder controle geplaatst van een ComK-afhankelijke promoter, dus met een K-box. Wanneer ComK transcriptie activeert door te binden aan deze box, wordt er een RNA-kopie van het *lacZ*-gen gemaakt. Dit gen codeert voor het eiwit betagalactosidase dat een kleurloze stof om kan zetten in een gele. De mate van geelkleuring wordt bepaald door de hoeveelheid aanwezig betagalactosidase en hangt af van de mate van transcriptie van *lacZ* onder controle van ComK. Uiteindelijk is de geelkleuring dus een indicatie voor de functionaliteit van ComK in transcriptie-activering.
Met behulp van deze experimenten is aangetoond dat een ComK-variant die 25 aminozuren van het einde van het eiwit mist, geen transcriptie meer kan activeren, maar nog wel DNA kan binden. Gel-retardaties lieten echter een kleiner verschil in loopsnelheid

zien tussen het ongebonden en het ComK-gebonden DNA dan met het originele ComK-eiwit. Dit duidt erop dat er minder ComK-eiwitten aan het DNA gebonden zijn en suggereert dat de mutant als dimeer bindt in plaats van als tetrameer. Hieruit kan geconcludeerd worden dat het tetramerisatiedomein van ComK zich aan het uiteinde van het eiwit bevindt.
Een andere ComK-mutant, die negen aminozuren vanaf het begin mist, is verrassenderwijs actiever in transcriptie-activering, hoewel er minder eiwit geproduceerd wordt. Waarschijnlijk speelt het eerste stuk van het eiwit een rol bij het verhogen van de productie en/of stabiliteit van ComK, maar gaat de aanwezigheid van dit gedeelte normaal gesproken ten koste van maximale transcriptie-activering. Het zou kunnen dat de lengte van deze regio een compromis vormt tussen hoge eiwitproductie en/of –stabiliteit en hoge transcriptie-activering. De exacte rol van deze regio is echter nog onduidelijk en opheldering hiervan vergt meer onderzoek.

Cruciale baseparen voor de activiteit van een K-box

Voor veel DNA-bindende eiwitten is de herkenningssequentie gedefinieerd op basis van overeenkomsten in de sequenties die voorkomen voor de start van de verschillende genen die onder controle van het eiwit staan. De gedefinieerde sequentie wordt de **consensus sequentie** genoemd en geeft voor elke positie in de herkenningssequentie aan welke base het meeste voorkomt. Voor ComK is bijvoorbeeld bekend dat een K-box bestaat uit twee van elkaar gescheiden AT-boxen met de sequentie AAAA-$(N)_5$-TTTT.
In het verleden is aangetoond dat K-boxen met maximaal drie afwijkingen van de consensus sequentie nog functioneel kunnen zijn. In het totale DNA van *B. subtilis* komen ruim 1000 K-boxen voor, die aan deze criteria voldoen. Ongeveer dertig procent van deze K-boxen ligt voor de promoter van een gen en zou dus betrokken kunnen zijn bij transcriptie-activering. Er zijn echter maar twee K-boxen die perfect overeenkomen met de consensus sequentie. In totaal wordt slechts acht procent van de mogelijk actieve K-boxen daadwerkelijk gereguleerd door ComK. Dit suggereert dat er, ondanks de grote natuurlijke variatie, kennelijk ook afwijkingen voorkomen die niet toegestaan zijn en die leiden tot inactieve K-boxen. Dit gegeven is als startpunt gebruikt voor een onderzoek naar cruciale posities in de consensus sequentie van

een K-box voor transcriptie-activering door ComK. Door systematisch elke adenine of thymine base in één van beide AT-boxen te vervangen door een guanine, is het effect onderzocht dat één afwijking heeft op activiteit van de K-box. Voor elke gemuteerde K-box zijn ComK-binding en transcriptie-activering bepaald en vergeleken met binding en activering van een perfecte K-box. Het opvallendste was de vermindering tot slechts 30% van het controle-niveau, wanneer de tweede T (T_2) in een AT-box vervangen werd door een G. Wanneer ook de tweede T in de andere AT-box vervangen werd door een G, was er nauwelijks nog ComK-binding en transcriptie-activering detecteerbaar (figuur 7).



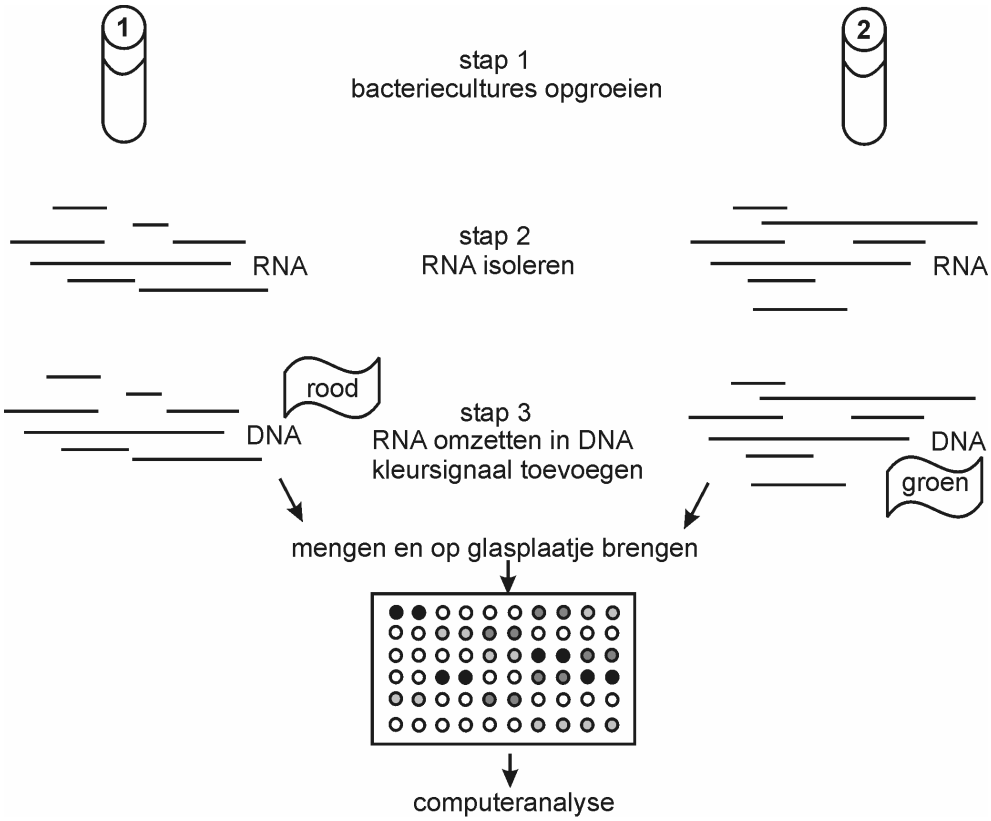
Figuur 7. Overzicht van de belangrijkste, cruciale posities in een K-box
De pijltjes geven de gemuteerde posities aan. Het percentage geeft de activiteit van de K-box ten opzichte van een perfecte K-box aan.

De experimentele resultaten werden ondersteund door een computeranalyse van alle bekende actieve K-boxen van *B. subtilis*. In totaal zijn dit er 88 en in geen enkele actieve K-box werd een mutatie aangetroffen in de T_2-posities van beide AT-boxen tegelijk. Bij inactieve K-boxen komen dubbele T_2-mutaties wel voor. Dit onderstreept dat K-boxen met dergelijke mutaties inderdaad niet meer actief zijn. De reden voor de inactiviteit van de T_2-gemuteerde K-boxen is tot dusver nog onduidelijk, maar het is zeer waarschijnlijk dat de T_2-positie een belangrijke rol speelt bij de binding van ComK aan de K-box en dat ComK de sterkste interactie met het DNA heeft op deze positie.

Introductie van ComK in een andere bacterie

De mogelijkheden van ComK om gentranscriptie te reguleren, zijn verder onderzocht in *Lactococcus lactis*, een melkzuurbacterie die van groot belang is voor de zuivelindustrie. Het genoom van *L. lactis* is ongeveer anderhalf keer zo klein als dat van *B. subtilis*,

maar het is eveneens zeer A/T-rijk. Een interessante observatie is dat het DNA van *L. lactis* veel K-boxen bevat en er dus in potentie veel regulatie door ComK mogelijk is. *L. lactis* heeft echter zelf geen *comK*-gen en dus geen ComK-eiwit. Zou ComK van *B. subtilis* in staat zijn om transcriptie te reguleren door te binden aan de K-boxen in *L. lactis*? Om deze vraag te kunnen beantwoorden, werd *comK* uit *B. subtilis* geïntroduceerd in *L. lactis*. Op basis van dit ingebrachte gen, kan een *L. lactis* cel *B. subtilis* ComK produceren. De functionaliteit van dit ComK werd bevestigd met een betagalactosidase-experiment, waarbij ComK in *L. lactis* in staat bleek om transcriptie te activeren van *lacZ* onder controle van een ComK-afhankelijke promoter, uit *B. subtilis*, met een K-box.
Transcriptie-regulatie door ComK vanaf K-boxen in *L. lactis* kan aangetoond worden met behulp van een DNA-microarray analyse (figuur 8). Deze methode vergelijkt het transcriptieniveau van alle genen in twee situaties, in dit geval *L. lactis* met of zonder ComK. De twee bacteriecultures worden onder gelijke omstandigheden opgegroeid tot er voldoende cellen aanwezig zijn. Uit deze cellen wordt RNA geïsoleerd. Er is alleen een RNA-afschrift aanwezig van genen waarvan op dat moment transcriptie plaatsvond. Om transcriptie aan te tonen, wordt van elk aanwezig RNA weer DNA gemaakt, waarbij een kleursignaal aan het DNA gekoppeld wordt, bijvoorbeeld rood voor *L. lactis* met ComK en groen voor *L. lactis* zonder ComK. Het DNA van beide situaties wordt samengevoegd en op een glasplaatje gebracht waaraan DNA gekoppeld zit van alle genen van *L. lactis*. Het toegevoegde DNA herkent het al aanwezige DNA van zijn eigen gen en bindt daar aan. Een computeranalyse bepaalt daarna per gen of er meer rood of meer groen DNA gebonden is. De kleur van het signaal is een indicatie voor de verschillen in transcriptie tussen beide situaties. Rood betekent dat er van een gen meer transcriptie is met ComK, dus dat de aanwezigheid van ComK transcriptie van dit gen stimuleert, terwijl groen aangeeft dat de aanwezigheid van ComK de transcriptie van een gen remt. Er zijn nog twee andere signalen mogelijk voor elk gen: geen signaal, als er in beide situaties geen transcriptie is of een geel signaal, als er in beide situaties evenveel transcriptie van het gen plaatsvindt.
Door gebruik te maken van de DNA-microarray-techniek, zijn de veranderingen in het transcriptieprofiel van *L. lactis* als gevolg van de productie van ComK in kaart gebracht. ComK beïnvloedde de transcriptie van 7% van alle genen.



Figuur 8. Het principe van een DNA-microarray
De eerste drie stappen worden voor culture 1 en 2 afzonderlijk uitgevoerd. Daarna wordt het gekleurde DNA samengevoegd en op een glasplaatje gebracht. Elk rondje op het glasplaatje bevat DNA van één gen van het *L. lactis* genoom. Alle rondjes tezamen bevatten het totale genoom. De kleuren geven verschil in transcriptie van een gen aan. Wit: in beide bacteriecultures geen transcriptie; lichtgrijs: in culture 1 meer transcriptie dan in 2; zwart: in culture 2 meer transcriptie dan in 1; donkergrijs: in beide cultures evenveel transcriptie.

Slechts 12% van de 89 geactiveerde genen en 27% van de 114 geremde genen werden voorafgegaan door een K-box. Voor deze groep genen kan de verandering in transcriptie een direct gevolg zijn van ComK-binding aan de K-box. Het grote aantal geremde genen met een K-box is opvallend, aangezien ComK in *B. subtilis* eigenlijk uitsluitend als activator werkt. Een mogelijke verklaring voor de remming van transcriptie is dat ComK aan de beschikbare K-boxen bindt en de normale transcriptie-regulatie verstoort, bij-

voorbeeld door RNAP-binding te blokkeren. Om deze verklaring te toetsen, is een tweede DNA-microarray studie uitgevoerd met de eerder beschreven ComK-variant, die geen transcriptie meer kan activeren, maar nog wel DNA kan binden. De geobserveerde effecten kunnen in dit geval alleen het gevolg zijn van binding en/of aanwezigheid van ComK. De analyse toonde veel overeenkomsten aan tussen de effecten van oorspronkelijk en mutant ComK op het transcriptieprofiel van *L. lactis*. Hieruit kan geconcludeerd worden dat de directe effecten van ComK voornamelijk veroorzaakt zijn door binding aan de K-boxen.
De meerderheid van de gereguleerde genen bevat echter geen K-box. De transcriptie van deze genen wordt dus op indirecte wijze beïnvloed door ComK. De cel kan bijvoorbeeld last hebben van de productie en/of aanwezigheid van ComK en kan hierdoor minder hard groeien. Als reactie kan de transcriptie van genen die betrokken zijn bij de opname van voedingsstoffen geremd worden, omdat de cel deze voedingsstoffen onder de gegeven omstandigheden toch niet volledig kan benutten. De achtergrond van de indirecte effecten kan aangegeven worden door de gereguleerde genen te groeperen op grond van hun functie. Een dergelijke analyse in dit geval, toonde aan dat de geremde genen inderdaad vooral betrokken zijn bij opname en gebruik van voedingsstoffen. De geactiveerde genen zijn vooral betrokken bij eiwitsynthese. Dit zou verklaard kunnen worden doordat ComK de transcriptie van genen remt, terwijl de bijbehorende eiwitten nog wel nodig zijn. De cel probeert dan om op basis van het nog aanwezige RNA, toch veel eiwit te maken en zo aan de vraag te voldoen.
Het interessante van deze studie is dat aangetoond is dat een regulator in de ene bacterie niet automatisch hetzelfde werkt als in een andere. In *B. subtilis* is ComK eigenlijk uitsluitend een activator, terwijl ComK-productie in *L. lactis* juist in repressie op grote schaal resulteert. Veel van de werking van een eiwit hangt dus ook af van de omgeving waarin het zich bevindt.

Algemene conclusies en suggesties voor verder onderzoek

De belangrijkste resultaten van het hier beschreven promotieonderzoek zijn de opheldering van de rol van ComK in het transcriptie-activeringsmechanisme, de identificatie van de regio van het ComK-eiwit die betrokken is bij tetramerisatie en de ontdek-

king dat de tweede thymine base in een AT-box van cruciaal belang is voor ComK-binding. Hoewel er natuurlijk altijd meer vragen te beantwoorden zijn, is het onderzoek naar de rol van ComK in transcriptie-activering hiermee afgesloten. Voor een compleet beeld van de functionele domeinen in ComK en de cruciale posities in een K-box is echter nog meer onderzoek nodig. Het zou natuurlijk zeer praktisch zijn wanneer je het ComK-eiwit gewoon zou kunnen zien. In dat geval zou duidelijk worden welke delen van het eiwit betrokken zijn bij een bepaalde functie. Wanneer je ComK gebonden aan DNA zou bekijken, zou je bijvoorbeeld het DNA-bindingsdomein van ComK aan kunnen wijzen, evenals de belangrijkste posities in de K-box. Normaal gesproken zijn eiwitten echter te klein om tot in detail te kunnen bekijken. Maar er bestaat een techniek waarmee eiwitten en ook eiwit-DNA-complexen zichtbaar gemaakt kunnen worden. Deze techniek wordt eiwitkristallisatie genoemd en maakt gebruik van de afbuiging van röntgenstraling door eiwitten en DNA. Dit levert een bepaald patroon op dat omgezet kan worden in een drie-dimensionaal beeld van het eiwit of het eiwit-DNA-complex. In de praktijk zijn deze studies vaak erg lastig en moeten er veel pogingen, onder verschillende condities ondernomen worden om uiteindelijk een stabiel eiwit of eiwit-DNA-complex te verkrijgen. Wanneer een complex echter eenmaal gevormd is, zijn veel details van het eiwit en het DNA zichtbaar. Een dergelijk beeld kan helpen om te bekijken welke aminozuren van het eiwit een interactie aangaan bij de vorming van een dimeer of tetrameer of welke basen van het DNA van belang zijn voor de eiwit-DNA-interacties. De verwachting is dat een kristallisatie-studie van een ComK-DNA-complex kan helpen bij het ophelderen van de precieze rol van de T_2-positie van een AT-box in ComK-binding. Verder kan de studie gebruikt worden om te bekijken welke aminozuren van de uiteinden van ComK betrokken zijn bij tetramerisatie. Er is een begin gemaakt met een kristallisatie-studie van een ComK-DNA-complex. Tot dusver heeft deze studie echter nog geen resultaten opgeleverd. Als alternatief is het natuurlijk ook mogelijk om meer eiwitmutanten te maken en deze te testen op bijvoorbeeld DNA-binding, dimerisatie en interacties met andere eiwitten, maar een succesvolle kristallisatie-studie zou een completer beeld kunnen geven van de verschillende functionele domeinen in ComK. Het is daarom aan te raden om de aandacht vooral te richten op het verkrijgen van goede ComK- en/of ComK-DNA-kristallen.

Nawoord

Iedereen bedankt!!

Op mijn babykamer hingen vroeger twee posters met tekeningen van Holly Hobbie. Op de ene poster stond: "Begin de morgen zonder zorgen" en op de andere stond: "Samen ben je niet alleen". Met deze uitspraken ben ik dus opgegroeid. Het eerste is niet altijd gemakkelijk, zeker niet voor een promovendus. De waarde van de tweede uitspraak is maar weer eens onderstreept door alle collega's van Moleculaire Genetica en de mensen buiten het lab, die mij tijdens mijn AIO-tijd gesteund hebben. En als er één ding van belang is voor een AIO, is het wel de steun van anderen. In het kort kan ik daarover zeggen: die was fantastisch! Iedereen heel hartelijk bedankt!

Natuurlijk wil ik enkele mensen apart bedanken. Ten eerste mijn promotor, Oscar Kuipers. Oscar, jouw enthousiasme en je positieve begeleiding heb ik altijd zeer gewaardeerd. Hoewel het onderzoek in het begin vaak moeizaam ging, hield je er vertrouwen in en, nog belangrijker, sprak je dat vertrouwen ook naar mij toe uit. Dat heeft me veel steun gegeven en zorgde ervoor dat ik het zelf ook weer positiever inzag. Na een bespreking ging ik vaak weer met hernieuwde energie en moed aan de slag. Bedankt voor alles! Ten tweede mijn co-promotor, Leendert Hamoen. Leendert, ik heb veel van je geleerd. Een betere leermeester voor eiwitzuiveringen, retardaties en transcripties kon ik me niet wensen. Ook toen je in Engeland zat, ben je altijd bij het onderzoek betrokken gebleven. Een begeleider op afstand was hier zeker geen probleem! Bedankt voor je betrokkenheid en je enthousiasme voor het competentie-onderzoek! Ik wil graag van de gelegenheid gebruik maken om ook Sierd Bron te bedanken. Sierd, hoewel je niet direct betrokken was bij mijn onderzoek, heb je me regelmatig geholpen. Zelfs wanneer je het erg druk had, maakte je toch even tijd vrij. Bedankt voor alles!

Op deze plaats wil ik ook de leden van de leescommissie, Prof. Dr. Jan Maarten van Dijl, Prof. Dr. Peter van Haastert en Prof. Dr. Ida van der Klei bedanken voor hun vlotte en gedegen beoordeling van het manuscript.

Een belangrijke bijdrage aan een geslaagde promotietijd wordt geleverd door fijne collega's en een plezierige werksfeer. En die vind je bij MolGen! Mijn eerste echte kennismaking met de groep was toen ik als student bij Rob Meima werkte. Rob, bedankt voor je plezierige en positieve begeleiding! Deze ervaring heeft zeker

een rol gespeeld bij mijn keuze om in wetenschappelijk onderzoek door te gaan. Toen ik later als AIO werd aangenomen, zag ik het werken bij MolGen dan ook met vertrouwen tegemoet: het onderzoek zal niet altijd meevallen, maar op de groep kan je rekenen. En dat is zeker uitgekomen. Ik wil dan ook alle (ex-)collega's bedanken, in willekeurige volgorde, Robèr, Andrzej, Harma, Aäron, Jetta, Siger, Anton, Elise, Chris, Evert-Jan, Wietske, Jan-Willem, Naomi, Tomas, Robyn, Maarten, Richard, Caroline, Dinne, Girbe, Patricia, Rense, Joanna, Morris, Cordula, Kees, Wiep Klaas, Anne H., Reindert, Olivera, Sacha, Michiel, Holger, Helga, Anne de J., Esther, Hein, Aldert, Anja, Jan J., Henk, Ulrike, Rasmus, Mariska, Jan K., Erwin, Arno, Emmo, Nathalie en alle buitenlandse gasten, studenten en mensen die ik onverhoopt vergeten ben. Ik heb het altijd enorm gewaardeerd dat iedereen bereid is om elkaar te helpen en dat mensen open staan voor vragen en suggesties van anderen. In zo'n omgeving is het prettig werken!
Een ander aspect dat het werken bij MolGen zo prettig maakt, zijn de luxe arbeidsomstandigheden met betrekking tot bestellingen, glaswerk en medium. Arie, Peter en Mozes, jullie nemen ons veel werk uit handen! Bedankt voor deze perfecte organisatie! Ook op het secretariaat is alles prima geregeld. Emma, bedankt voor je uitstekende werk! Margriet en Mirelle, bedankt voor jullie inzet! Op deze plaats wil ik ook Peter Wiersema, Cees Vermeulen en Arjo Bunskoeke bedanken voor de goede zorgen op het isotopenlab.
Enkele collega's wil ik nader noemen, te beginnen bij mijn zaalgenoten. Erwin, jij was vaak mijn eerste aanspreekpunt bij vragen of frustraties over het onderzoek. Bedankt voor al je hulp, je adviezen en de gezellige praatjes in ons kleine schrijfhok! Evert-Jan, Dinne en Morris, ook met jullie heb ik een prima tijd gehad. Wiep Klaas en Jan-Willem (en Reindert, ook al ben je geen zaalgenoot), jullie gedrevenheid en creatieve oplossingen zorgden altijd voor een positieve werksfeer! Aske, hoewel jij al weg was toen ik als AIO begon en we dus nooit samengewerkt hebben, is hoofdstuk twee toch het resultaat van ons gezamenlijk werk. Bedankt voor de goede basis die jij al gelegd had!
Een gedeelte van mijn onderzoek heb ik uitgevoerd in *Lactococcus lactis* in plaats van in *Bacillus subtilis*. Anton en Rasmus, jullie hebben mij geweldig geholpen om snel vertrouwd te raken met *L. lactis*! Naomi, bedankt voor je hulp bij de arrays! Natuurlijk wil ik ook mijn studenten, Jan, Ewa en Mike, hartelijk bedanken. Ik

vond studenten-begeleiding altijd een van de leukste onderdelen van mijn werk en heb de samenwerking met jullie als zeer plezierig ervaren. Jan, bedankt voor je uitstekende inzet! My two Polish students, Ewa and Mike, thank you for the work you did.
Goede herinneringen heb ik ook aan de samenwerking met Fabrizia Fusetti en Andy-Mark Thunnissen van eiwitkristallografie. Besprekingen met jullie vond ik altijd erg interessant. Bedankt voor de leuke samenwerking!
Mijn paranimfen, Chris en Esther, verdienen een speciale plaats in dit nawoord. Chris, in hoofdletters BEDANKT dat je me geholpen hebt om mijn impasse met de RNAP-retardaties te doorbreken! Naast de wetenschappelijke samenwerking kijk ik ook met plezier terug op de prettige tijd op het lab, de ontspannen gesprekjes en de onderlinge geintjes over voetbal. Esther, ik vond het altijd erg leuk om ’s ochtends vroeg, ’s middags bij de lunch en soms bij de theepauze even te kletsen. Bedankt voor alle gezelligheid! Chris en Esther, ik ben erg blij dat jullie mijn paranimfen zijn en me willen vergezellen op mijn promotiedag!
Hoewel familie, vrienden en kennissen geen directe bijdrage leveren aan de wetenschappelijke inhoud van je proefschrift, kunnen ze je promotietijd wel gemakkelijker maken. Iedereen heel hartelijk bedankt voor de belangstelling en de geboden ontspanning!
Aan het eind van dit dankwoord wil ik nog enkele mensen noemen die voor mij zeer speciaal zijn, mijn oma en mijn ouders. Oma, bedankt voor al je belangstelling! Ik vind het fantastisch dat je er op mijn promotiedag bij kan zijn! Peter en Jeannette, jullie zijn er altijd voor mij! Bedankt voor alles, voor alle steun, liefde en vertrouwen! Door jullie geloof in een goede afloop van mijn promotie-onderzoek, ben ik er zelf ook in blijven geloven!